

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:02:58 ; Search time 1188 Seconds  
(without alignments)  
5879.355 Million cell updates/sec

Title: US-09-980-277-1

Perfect score: 240

Sequence: 1 gatcgtacgcgacacacg.....cgaattatattttctt 240

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_lm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_in:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vfl:\*  
38: em\_sy:\*  
39: em\_htg\_hum:\*  
40: em\_htg\_mus:\*  
41: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44.2	18.4	121	3	SCMRPTAA
2	40.8	17.0	69028	9	AC087493
3	40.8	17.0	150794	9	AC025211
4	40.8	17.0	186360	9	AC007982
5	40.6	16.9	218922	2	AC008106
6	40.6	16.9	171744	2	AC095286
7	39.8	16.6	57075	9	AC080073
8	39.2	16.3	180104	9	DJ526N18
9	38.4	16.0	166434	2	AC109963
10	38.4	16.0	192153	2	AC095187
11	38	15.8	32940	3	AF022972
12	37.6	15.7	114473	2	AC113374
13	37.6	15.7	177720	2	AC007514
14	37.6	15.7	190481	2	AC125626
15	37.4	15.6	220181	2	AC094816
16	37.4	15.6	222004	2	AC023584
17	37.2	15.5	68133	2	AC100012
18	37.2	15.5	165733	2	AC099146
19	37	15.4	168142	2	AC102140
20	37	15.4	173987	2	AL450337
21	36.8	15.3	41101	8	SPAC144
22	36.6	15.2	6134	6	XZ51318
23	36.6	15.2	101213	2	AC113851
24	36.6	15.2	14680	2	AC095815
25	36.6	15.2	154960	2	AC011948
26	36.2	15.1	153735	9	AC004029
27	36	15.0	13416	1	AE007832
28	36	15.0	144884	2	AC096990
29	36	15.0	169688	2	AC099745
30	36	15.0	173786	2	AC107582
31	36	15.0	191720	2	AC109982
32	36	15.0	225407	2	AC124428
33	35.8	14.9	62418	2	AC104025
34	35.8	14.9	174186	2	AC025252
35	35.8	14.9	174293	2	AC068787
36	35.8	14.9	196698	9	CNS05RET
37	35.8	14.9	224817	10	AL663048
38	35.8	14.9	347726	2	AC098521
39	35.6	14.8	148756	2	AC119598
40	35.6	14.8	170659	2	AC112694
41	35.6	14.8	196410	9	AC027544
42	35.6	14.8	198565	2	AC016750
43	35.6	14.8	255839	2	AC105842
44	35.4	14.8	8623	3	AY082576
45	35.4	14.8	28759	9	AC062027

## ALIGNMENTS

RESULT 1  
LOCUS SCMRPTAA  
DEFINITION S. mansoni tandem repeat units.  
ACCESSION M61098  
VERSION M61098.1 GI:161064  
KEYWORDS  
SOURCE  
ORGANISM  
S. mansoni (strain Egyptian) DNA, clone pSm1-7.  
Schistosoma mansoni  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE  
AUTHORS  
TITLE  
1 (bases 1 to 121)  
Hammerling, J., Turetski, T., Kapeller, I. and Deresiewicz, R.  
Highly repeated short DNA sequences in the genome of Schistosoma  
mansoni recognized by a species-specific probe

1

```
repeat_region      815. .8321 /rpt_family="ALUSx"
```

	repeat_region	27207 - 27771	/pt_family="LIMC"	
Query Match		17.0%	Score 40.8;	DB 9; Length 69028;
Best Local Similarity		49.5%;	Pred. No. 0.8;	
Matches 105; Conservative		0;	Mismatches 107;	Indels 0; Gaps 0;
QY	22	TCATGGAAGAACGTTGTAATCAGACTTATGCGTGTGGCAAGATCTTTAGCAACATAC	81	
Db	11615	TATATTAACACCTTAAAGAGATTTATATTTGTACAAATGTATATCATTTAAATTA	11674	
QY	82	TCGGAACCACTGGACGAGATTTTATGATGTTTGTATAGAGCTTTGGTCACTGCCGA	141	
Db	11675	TCGATATTCAGTTTAAACCTCAATATAAAATTTTCAATTTATATTAACCTATATGCACTTTAAA	11734	
QY	142	AAATATCTACAAACAAATATGATTATTTTCGAGACGCTGGCGTTAATATAAACAGAAC	201	
Db	11735	TGTATATCGCAGATTAAGAAGTCAATGATATAGTATATCAACCTTATATAAGAGATGAAA	11794	
QY	202	TATATAAGCTCTCGCACCGCGCAATTAATTT	233	
Db	11795	AGATTAAGTAAACCTCTTAACACCTTAATATAT	11826	

RESULT 3	AC025211	150794 bp	DNA	linear	PR1 08-SBP-2001
LOCUS	AC025211	150794 bp	DNA	linear	PR1 08-SBP-2001
DEFINITION	Homosapiens chromosome , clone RP11-220C2, complete sequence.				
ACCESSION	AC025211				
VERSION	AC025211.8	GI:15528921			
KEYWORDS	HTG.				
ORGANISM	Homosapiens				
	Homosapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 150794)				
JOURNAL	Homosapiens chromosome, clone RP11-220C2				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 150794)				
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,				
	Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,				
	Boguslavsky,L., Bonkagalter,B., Brown,A., Burkett,G.,				
	Campyano,A., Castle,A., Chospel,Y., Colangelo,M., Collins,S.,				
	Collimore,A., Cooke,P., Dekrelano,K., Dewar,K., Diaz,T.S.,				
	Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,				
	Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,				
	Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,				
	Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karats,A.,				
	Klein,J., Larcoque,K., Lamazares,R., Landers,T., Lebocky,J.,				
	Levine,K., Lien,C., Liu,G., Locke,K., MacDonald,P., Margus,N.,				
	McCarthy,M., McKean,P., McGuck,A., McKernan,K., McPheters,R.,				
	Meldrum,D., Meneus,J., Mihova,T., Miranda,C., Mlenya,V., Morrow,J.,				
	Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,				
	O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,				
	Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,				
	Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,				
	Stagge-Thomson,N., Stojanovic,N., Sudramanian,A., Talamas,J.,				
	Teisberg,S., Theodore,J., Tittell,A., Travers,M., Tiggilio,J.,				
	Vassiliev,B., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,				
	Young,G., Zainoum,J., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-MAR-2000) Whitehead Institute/MIT Center for Genome				
REFERENCE	Research, 320 Garden Street, Cambridge, MA 02141, USA				
AUTHORS	3 (bases 1 to 150794)				
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,				
	Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bonkagalter,B.,				
	Brown,A., Camarata,J., Campyano,A., Chan,J., Chazaro,B.,				
	Chospel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A.,				
	Cooke,P., Dekrelano,K., Dewar,K., Diaz,J.S., Dodge,S., Fero,S.,				
	Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,				
	Ginde,S., Gind,S., Goyette,M., Graham,L., Grand-pierre,N.,				
	Hagos,B., Heaford,A., Horton,L., Hutme,W., Iliev,I., Johnson,R.,				

TITLE  
JOURNAL

## COMMENT

Jones, C., Kamat, A., Karatas, A., Kells, C., Laboquie, K.,  
Iamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Margus, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meltrin, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Rella, R., Rhinck, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,  
Seaman, S., Severy, P., Spencer, B., Strange-Thomson, S., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testfay, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (08-SEP-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 8, 2001 this sequence version replaced g1:15431159.  
All repeats were identified using RepeatMasker:  
Smith, A. F. A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L7840  
Center clone name: 220\_C2  
-----

FEATURES  
SP6 end overlaps AC002119 [WTCGR L141] by 43891 bp. We will submit  
150793 bp of the overlap.  
Location/Qualifiers

1. 150794  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-220C2"  
/clone\_id="RP11-220C2-11 Human Male BAC"  
complement(1..23)  
/rpt\_family="L1ME4A"  
6..23  
/note="<30 qual SINGL region"  
24..328  
/rpt\_family="AlusC"  
complement(329..425)  
/rpt\_family="L1ME4A"  
complement(599..828)  
/rpt\_family="Aluub"  
complement(987..1086)  
/rpt\_family="L1ME4A"  
complement(1087..1381)  
/rpt\_family="Aluub"  
complement(1382..1430)  
/rpt\_family="L1ME4A"  
complement(1532..1780)  
/rpt\_family="L1ME4A"  
complement(1816..2132)  
/rpt\_family="Alusx"  
complement(3188..3312)  
/rpt\_family="L2"  
complement(3624..3919)  
/rpt\_family="Alusx"  
complement(4349..4640)  
/rpt\_family="Alusx"  
complement(4861..5224)  
/rpt\_family="L2"  
5818..5936  
/rpt\_family="L2"  
complement(5992..6226)  
/rpt\_family="MIR"  
7173..7280  
/rpt\_family="MER94"

repeat\_region 7486..7679  
/rpt\_family="L2"  
complement(8988..9267)  
/rpt\_family="AluX"  
10907..11697  
/rpt\_family="L2"  
complement(11992..12132)  
/rpt\_family="HAL1"  
complement(12133..12440)  
/rpt\_family="Aluub"  
complement(12441..12462)  
/rpt\_family="HAL1"  
complement(12463..12767)  
/rpt\_family="AluY"  
complement(12768..12898)  
13422..13629  
/rpt\_family="HAL1"  
15268..15291  
/rpt\_family="MIR"  
/rpt\_family="CTGAA)n"  
complement(15292..15580)  
/rpt\_family="Aluub"  
15584..15883  
/rpt\_family="Alusx"  
complement(15851..15659)  
/note="<30 qual SINGL region"  
complement(15684..15701)  
/note="<30 qual SINGL region"  
complement(15859..15865)  
/note="<30 qual SINGL region"  
complement(15893..16047)  
/rpt\_family="Aluub"  
complement(16350..16654)  
17539..17654  
/rpt\_family="AluY"  
/rpt\_family="L2"  
complement(17671..17984)  
/rpt\_family="Aluub"  
complement(18115..18681)  
/rpt\_family="L1ME1"  
complement(18783..19023)  
/rpt\_family="L1ME1"  
complement(19024..19321)  
/rpt\_family="Alusg"  
complement(19322..20092)  
20098..20194  
/rpt\_family="L1ME1"  
/rpt\_family="CT-rich"  
complement(20195..20356)  
/rpt\_family="Aluub"  
20357..20395  
/rpt\_family="TGN"  
complement(20396..20690)  
/rpt\_family="Alusx"  
complement(20691..20830)  
/rpt\_family="Aluub"  
complement(20831..22342)  
/rpt\_family="L1ME1"  
complement(21587..21602)  
/note="<30 qual SINGL region"  
complement(21644..21658)  
/note="<30 qual SINGL region"  
complement(22343..22535)  
/rpt\_family="Aluub"  
complement(22629..22718)  
/rpt\_family="AluY/FLAM"  
complement(23036..23116)  
/rpt\_family="L1MB2"  
complement(23117..23399)  
/rpt\_family="Aluub"  
23399..23399  
/rpt\_family="Aluub"  
complement(23400..23422)  
repeat\_region

```

repeat_region      /rpt_family="L1MB2"
                    complement(23472..23561)
repeat_region      /rpt_family="MIR"
                    complement(24063..24374)
repeat_region      /rpt_family="AluX"
                    complement(24989..25192)
                    /rpt_family="L3"

```

Query Match 17.0%; Score 40.8; DB 9; Length 150794;

Best Local Similarity 49.5%; Pred. No. 0.76; Indels 0; Gaps 0;

Matches 105; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

QY 22 TCTATGAAATCGTGTATCTAGCTAGCTGGTGGGCAAGATCTTTAGCACTAC 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148581 TATATACACTCTTATATAGATGATTTATTTGTAAGATTTATTCATTCGAATTA 148640
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 82 TCCGAACCCAGTGGACGATTTTATGATTTTGTATGAGGCTTTGGTACCTGCTTA 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148641 TCTATATTCAGTTCAATCCATAATAATTTGTAATTAATTAACCTTATGCTATAAA 148700
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 AATATCTACAAACAAATATGATTTTTCGAGACCGTGGCGTTATATTAACAAGAAC 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148701 TGTATATCTGCAAGATTAAGGTATATGCTATATTAACACCTTATTAAGAAGTAA 148760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 202 TATATAACGCTTCGCAACCCGCAATTAATTT 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148761 AGATTAATCTAACCCCTTAACACATTAATAT 148792
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 4
LOCUS AC007982/c 186360 bp DNA linear PRI 31-MAY-2002
DEFINITION Homo sapiens chromosome 17, clone RP11-44H5, complete sequence.
ACCESSION AC007982
VERSION AC007982.2 GI:21281552
KEYWORDS HGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 186360)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-44H5
Unpublished
2 (bases 1 to 186360)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Fuhe,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heatford,A., Horton,L., Howland,J.C., Jones,C., Kam,L.,
Karatas,A., Lehoczeky,J., Lieu,C., Locke,K., MacDonald,P.,
Margulis,N., McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J.,
Medrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torriella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (07-JUL-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 186360)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., FitzHugh,M., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A.,

```

```

Lamazares,R., Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., MacDonald,P., Major,J., Margulis,N.,
Mathews,C., McCarthy,M., McEwan,P., McKernan,K., Medrim,J.,
Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S.,
Schback,R., Seaman,S., Severy,N., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 31, 2002 this sequence version replaced gi:3391470.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1066
Center clone name: 44_H_5
-----

```

```

FEATURES
SOURCE
location/Qualifiers
1..186360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-44H5"
/clone.lib="RPCT-11 Human Male BAC"
496..501
/note="330 qual SINGL region."
505..648
/rpt_family="MER5B"
534..539
/note="330 qual SINGL region."
554..557
/note="330 qual SINGL region."
599..607
/note="330 qual SINGL region."
complement(923..1126)
/rpt_family="L1MB4A"
complement(1153..1446)
/rpt_family="L12"
complement(1490..1781)
/rpt_family="Alubd"
2152..2455
/rpt_family="Aluv"
complement(3160..3275)
/rpt_family="ML11L"
complement(3449..3711)
/rpt_family="ML11L"
3718..3940
/rpt_family="L3"
4189..4680
/rpt_family="L1MC4a"
4729..5262
/rpt_family="L1MC4"
5267..5430
/rpt_family="L3"
complement(5915..6213)
/rpt_family="Aluv"
complement(6302..6529)
/rpt_family="MIR"
complement(6701..7286)
/rpt_family="L2"

```

[illegible]

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTHR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 1900
Center clone name: 3008.D.3
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-Primer-amersham; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 197177 bases at least Q40
Consensus quality: 206822 bases at least Q30
Consensus quality: 211643 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 215622; sum-of-ctrls
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
* 1091 1190: contig of 1090 bp in length
* 1191 2690: contig of 1500 bp in length
* 2691 2790: gap of 100 bp
* 2791 4012: contig of 1222 bp in length
* 4013 4112: gap of 100 bp
* 4113 5123: contig of 1011 bp in length
* 5124 5223: gap of 100 bp
* 5224 6112: contig of 1089 bp in length
* 6113 6412: gap of 100 bp
* 6413 8474: contig of 2062 bp in length
* 8475 8574: gap of 100 bp
* 8575 10328: contig of 1754 bp in length
* 10329 10428: gap of 100 bp
* 10429 12116: contig of 1688 bp in length
* 12117 12216: gap of 100 bp
* 12217 14183: contig of 1967 bp in length
* 14184 14283: gap of 100 bp
* 14284 17330: contig of 3047 bp in length
* 17331 17430: gap of 100 bp
* 17431 21134: contig of 3704 bp in length
* 21135 21234: gap of 100 bp
* 21235 24428: contig of 3194 bp in length
* 24429 24528: gap of 100 bp
* 24529 27705: contig of 3177 bp in length
* 27706 27805: gap of 100 bp
* 27806 30106: contig of 2301 bp in length
* 30107 30206: gap of 100 bp
* 30207 35366: contig of 5160 bp in length
* 35367 35466: gap of 100 bp
* 35467 39634: contig of 4168 bp in length
* 39635 39734: gap of 100 bp
* 39735 43767: contig of 4033 bp in length
* 43768 43867: gap of 100 bp
* 43868 50104: contig of 6237 bp in length
* 50105 50204: gap of 100 bp
* 50205 54896: contig of 4692 bp in length
* 54897 54996: gap of 100 bp
* 54997 61677: contig of 6681 bp in length
* 61678 61777: gap of 100 bp
* 61778 68042: contig of 6265 bp in length
* 68043 68142: gap of 100 bp
* 68143 76003: contig of 7861 bp in length
* 76004 76103: gap of 100 bp
* 76104 83935: contig of 7832 bp in length

```

```

* 83936 84035: gap of 100 bp
* 84036 92731: contig of 8696 bp in length
* 92732 92831: gap of 100 bp
* 92832 102205: contig of 9374 bp in length
* 102206 102305: gap of 100 bp
* 102306 111555: contig of 9250 bp in length
* 111556 111655: gap of 100 bp
* 111656 120772: contig of 9117 bp in length
* 120773 120872: gap of 100 bp
* 120873 129869: contig of 8997 bp in length
* 129870 129969: gap of 100 bp
* 129970 143160: contig of 13191 bp in length
* 143161 143260: gap of 100 bp
* 143261 145441: contig of 11281 bp in length
* 145442 154641: gap of 100 bp
* 154642 170485: contig of 15844 bp in length
* 170486 170585: gap of 100 bp
* 170586 183973: contig of 13388 bp in length
* 183974 184073: gap of 100 bp
* 184074 200761: contig of 16688 bp in length
* 200762 200861: gap of 100 bp
* 200862 218922: contig of 18061 bp in length.
-----
FEATURES
source
1. 218922
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CTD-3008D3"
/clone_1lb="CTD Human BAC"
1. 1090
/misc_feature
/note="assembly-fragment"
1191. 2690
/misc_feature
/note="assembly-fragment"
2791. 4012
/misc_feature
/note="assembly-fragment"
4113. 5123
/misc_feature
/note="assembly-fragment"
5224. 6312
/misc_feature
/note="assembly-fragment"
6413. 8474
/misc_feature
/note="assembly-fragment"
8575. 10328
/misc_feature
/note="assembly-fragment"
10429. 12116
/misc_feature
/note="assembly-fragment"
12217. 14183
/misc_feature
/note="assembly-fragment"
14284. 17330
/misc_feature
/note="assembly-fragment"
17431. 21134
/misc_feature
/note="assembly-fragment"
21235. 24428
/misc_feature
/note="assembly-fragment"
24529. 27705
/misc_feature
/note="assembly-fragment"
27806. 30106
/misc_feature
/note="assembly-fragment"
30207. 35366
/misc_feature
/note="assembly-fragment"
35467. 39634
/misc_feature
/note="assembly-fragment"
39735. 43767
/misc_feature
/note="assembly-fragment"
43868. 50104
/misc_feature
/note="assembly-fragment"
50205. 54896
/misc_feature
/note="assembly-fragment"
54997. 61677
/misc_feature
/note="assembly-fragment"
61778. 68042
/misc_feature
/note="assembly-fragment"
68143. 76003
/misc_feature

```

```

misc_feature /note="assembly_fragment"
76104..83935
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
84036..92731
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
92832..102205
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
102306..111353
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
111656..120772
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
120873..129869
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
129970..143160
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
143261..151341
/note="assembly_fragment"

Query Match 17.0%; Score 40.8; DB 2; Length 218922;
Best Local Similarity 49.5%; Pred. No. 0.75;
Matches 105; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 22 TCGAAGAAACGCTGATGATGAGCTGCTGCGAGAGTCTTTAGCAGACATAC 81
Db 194649 TATATACCTCTTATATAGATGATTTATATGTAAGAAAGCTTATCAATTA 194708
QY 82 TCGAAGAAACGCTGAGCGATTTTATGATGTTGTTTAAAGCGCTTGTGACCTGCTTA 141
Db 194709 TCTATATTCAGTTCAATCCATTAATAAATAATTTGATTTAAATTAATGCTACTTAA 194768
QY 142 AAAATACATCAACAACAAATGATTTGCGAGCGCGCGCTTATATTAACAAGAAC 201
Db 194769 TGTATCTCGAGATTAAGATCTATGCTATAGCTTATATCAACCTTATTAAGAGTAA 194828
QY 202 TAATAAAGCTCTCGACCGCACTTATATTT 233
Db 194829 AGATAAATCAACCTTAACACACATATATAT 194860

RESULT 6
AC095286 171744 bp DNA linear HTG 11-JUL-2002
LOCUS Rattus norvegicus clone CH230-12C16, *** SEQUENCING IN PROGRESS
AC095286
VERSION AC095286.3 GI:21723052
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 171744)
Muzny D.M., Adams C., Adio-Ogunla B., Ali-Osman F.R., Allen C.,
Alshrocks S.L., Amaratunga H.C., Are J.R., Ayale M., Banks T.,
Barbarta J., Benton J., Blinze K., Blumhagen K., Bonin D.,
Bouck J., Bowie S., Brileva M., Brown E., Brown M., Bryant N.P.,
Bunay C., Burck P., Burkett C., Burrell K.L., Byrd N.C.,
Carton T.F., Carter M., Cavazos S.R., Chacko D., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dalbore S.R., David R.,
David M.L., Davis C., Davy-Carroll L., Dederich D.A.,
Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.B.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.U.,
Einhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez D.,
Hernandez O., Hodgson A., Hognes M., Holloway C., Hollins B.,
Hosni F., Howard S., Huber J., Hulik S., Hummel J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolyet S., Joudah S.,
Karlsson E., Kelly S., Khan U., King L., Korvah O., Kovar C.,

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 171744)
Worley K.C.
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 171744)
Worley K.C.
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GDUO
Center clone name: CH230-12C16
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 116672 bases at least Q40
Consensus quality: 123027 bases at least Q30
Consensus quality: 127822 bases at least Q20
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 68 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
1397 contig of 1396 bp in length
1487 contig of 1153 bp in length
2650 contig of unknown length
2749 contig of 1157 bp in length
3906 contig of unknown length
3907 contig of unknown length
4007 contig of 1010 bp in length
5016 contig of unknown length
5117 contig of 1213 bp in length
6328 contig of unknown length
6429 contig of 1425 bp in length
7854 contig of unknown length
7855 gap of unknown length

```

```

* 7955 9302: contig of 1348 bp in length
* 9303 9402: gap of unknown length
* 9403 10892: contig of 1490 bp in length
* 10893 10992: gap of unknown length
* 10993 12542: contig of 1550 bp in length
* 12543 12642: gap of unknown length
* 12643 13700: contig of 1058 bp in length
* 13701 13800: gap of unknown length
* 13801 14835: contig of 1035 bp in length
* 14836 14935: gap of unknown length
* 14936 16473: contig of 1538 bp in length
* 16474 16573: gap of unknown length
* 16574 18097: contig of 1524 bp in length
* 18098 18197: gap of unknown length
* 18198 19475: contig of 1279 bp in length
* 19477 19576: gap of unknown length
* 19577 20672: contig of 1096 bp in length
* 20673 20772: gap of unknown length
* 20773 22484: contig of 1712 bp in length
* 22485 22584: gap of unknown length
* 22585 24397: contig of 1813 bp in length
* 24398 24497: gap of unknown length
* 24498 26493: contig of 1996 bp in length
* 26494 26593: gap of unknown length
* 26594 28669: contig of 2076 bp in length
* 28670 28966: contig of 1127 bp in length
* 28967 29966: gap of unknown length
* 29967 31368: contig of 1372 bp in length
* 31369 31468: gap of unknown length
* 31469 33202: contig of 1734 bp in length
* 33203 33302: gap of unknown length
* 33303 35107: contig of 1805 bp in length
* 35108 35207: gap of unknown length
* 35208 37039: contig of 1832 bp in length
* 37040 37139: gap of unknown length
* 37140 39050: contig of 1911 bp in length
* 39051 39150: gap of unknown length
* 39151 40878: contig of 1728 bp in length
* 40879 40978: gap of unknown length
* 40979 43526: contig of 2548 bp in length
* 43527 43626: gap of unknown length
* 43627 44913: contig of 1287 bp in length
* 44914 45013: gap of unknown length
* 45014 46466: contig of 1453 bp in length
* 46467 46566: gap of unknown length
* 46567 48405: contig of 1839 bp in length
* 48406 48505: gap of unknown length
* 48506 50475: contig of 1970 bp in length
* 50476 50575: gap of unknown length
* 50576 51954: contig of 1379 bp in length
* 51955 52054: gap of unknown length
* 52055 54498: contig of 2444 bp in length
* 54499 54598: gap of unknown length
* 54599 57961: contig of 3363 bp in length
* 57962 58061: gap of unknown length
* 58062 59731: contig of 1670 bp in length
* 59732 59831: gap of unknown length
* 59832 61144: contig of 1313 bp in length
* 61145 61244: gap of unknown length
* 61245 62564: contig of 1320 bp in length
* 62565 62664: gap of unknown length
* 62665 64872: contig of 2208 bp in length
* 64873 64972: gap of unknown length
* 64973 67564: contig of 2552 bp in length
* 67565 67664: gap of unknown length
* 67665 69460: contig of 1796 bp in length
* 69461 69560: gap of unknown length
* 69561 72505: contig of 2945 bp in length
* 72506 72605: gap of unknown length
* 72606 76677: contig of 4072 bp in length
* 76678 76777: gap of unknown length
* 76778 79723: contig of 2946 bp in length

```

```

* 79724 79823: gap of unknown length
* 79824 82619: contig of 2796 bp in length
* 82620 82720: gap of unknown length
* 82720 85163: contig of 2444 bp in length
* 85164 85264: gap of unknown length
* 85264 88694: contig of 3431 bp in length
* 88695 88794: gap of unknown length
* 88795 91392: contig of 2598 bp in length
* 91393 91492: gap of unknown length
* 91493 94703: contig of 3211 bp in length
* 94704 94803: gap of unknown length
* 94804 97009: contig of 2206 bp in length
* 97010 97109: gap of unknown length
* 97110 99202: contig of 2093 bp in length
* 99203 99302: gap of unknown length
* 99303 101344: contig of 2042 bp in length
* 101345 101444: gap of unknown length
* 101445 104544: contig of 3100 bp in length
* 104545 104644: gap of unknown length
* 104645 108315: contig of 3671 bp in length
* 108316 108415: gap of unknown length
* 108416 111355: contig of 2940 bp in length

Query Match      16.9%; Score 40.6; DB 2; Length 171744;
Best Local Similarity 62.9%; Pred. No. 0.86;
Matches 61; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 141 AAAAACTACAAACAAATGATTTATTCGAGAGCGGCGCTTAATTAACACAGAA 200
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3376 AAAAAAAAAAAAAAAAAATTAATTTCTTTGGGGGCGCGGCGGAGAAAGAAAGAA 3435
QY 201 CTAAATAAGCTCTCGACCCCGCAATTAATTTGTT 237
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3436 AAAAAACAAGCTCCGCCGCCCGCENNTGTGTGTGCT 3472

RESULT 7
AC080073/c      AC080073      57075 bp      DNA      linear      PRI 09-JAN-2002
DEFINITION     Homo sapiens BAC clone RP11-321C7 from 7, complete sequence.
ACCESSION      AC080073
VERSION        AC080073.6 GI:15638817
KEYWORDS       HMG.
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 57075)
AUTHORS        Sulston, J.E. and Waterston, R.
TITLE          Toward a complete human genome sequence
JOURNAL        Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE        99063792
PUBMED         9847074
REFERENCE      2 (bases 1 to 57075)
AUTHORS        Trani, L., Kozlowicz, A. and Elliott, G.
TITLE          The sequence of Homo sapiens BAC clone RP11-321C7
JOURNAL        Unpublished (2001)
REFERENCE      3 (bases 1 to 57075)
AUTHORS        Waterston, R. H.
TITLE          Direct Submission
JOURNAL        Submitted (23-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      4 (bases 1 to 57075)
AUTHORS        Waterston, R. H.
TITLE          Direct Submission
JOURNAL        Submitted (18-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      5 (bases 1 to 57075)
AUTHORS        Waterston, R.
TITLE          Direct Submission
JOURNAL        Submitted (09-JAN-2002) Department of Genetics, Washington

```

**COMMENT**

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 18, 2001 this sequence version replaced gi:15487471.

----- Genome Center

Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)

## Summary Statistics

Center project name: H\_NH0321C07

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 2 Mapping Project (Eric D. Green, director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/dicr/grs/cmr/>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP01-1 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Feenberg, E., Tatematsu, M., Catanes, J. U. and de Jong, P. J. (1988). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 5:11-18. The clone may be obtained either from Research Genetics, Inc. (<http://www.rgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://baopac.med.buffalo.edu>).

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP4-736H5, 2000 bp overlap; the clone sequenced to the right is RP11-421N10, 2000 bp overlap. Actual start of this clone is at base position 109513 of RP4-736H5.

Polymorphisms have been indentified between AC005482 and AC080073.

## FEATURES

URS	location/Qualifiers
source	1. 57075
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="RP11-321C7"
	/clone_lib="RPC1-11"
	2. 68
repeat_region	/rpt_family="MR"
	69. 431
repeat_region	/rpt_family="MALR"
	500. .774
repeat_region	/rpt_family="L2"
	1485. 1589
repeat_region	/rpt_family="L2"
	2070. .2443
repeat_region	/rpt_family="MALR"
	3326. 3631
repeat_region	/rpt_family="ALU"
	3623. .3742
repeat_region	/rpt_family="MER1_type"
	3442. .3956

repeat_region	/rpt_family="MIR"	4038..4338
repeat_region	/rpt_family="Alu"	4342..4499
repeat_region	/rpt_family="Alu"	4498..4535
repeat_region	/rpt_family="(TA)n"	4534..4608
repeat_region	/rpt_family="ERV1"	4744..4780
repeat_region	/rpt_family="(TTTGG)n"	5396..5707
repeat_region	/rpt_family="Alu"	5798..6190
repeat_region	/rpt_family="ERV1"	6611..6786
repeat_region	/rpt_family="Alu"	6798..6847
repeat_region	/rpt_family="Alu"	7496..7594
repeat_region	/rpt_family="ERV1"	7591..7966
repeat_region	/rpt_family="ERV1"	8041..8340
repeat_region	/rpt_family="Alu"	8342..8738
repeat_region	/rpt_family="ERV1"	8657..8720
repeat_region	/rpt_family="AT-rich"	8799..9087
repeat_region	/rpt_family="Alu"	8949..8969
repeat_region	/rpt_family="AT-rich"	9154..9227
repeat_region	/rpt_family="L2"	9228..9848
repeat_region	/rpt_family="ERV1"	9849..9917
repeat_region	/rpt_family="L2"	10095..10828
repeat_region	/rpt_family="ERV1"	10955..11054
repeat_region	/rpt_family="L2"	11055..11355
repeat_region	/rpt_family="Alu"	11356..11490
repeat_region	/rpt_family="L2"	11832..12881
repeat_region	/rpt_family="ERV1"	13353..13652
repeat_region	/rpt_family="Alu"	13370..13397
repeat_region	/rpt_family="AT-rich"	13611..14247
repeat_region	/rpt_family="ERV1"	14248..14810
repeat_region	/rpt_family="ERV1"	14687..15449
repeat_region	/rpt_family="L1"	15504..16135
repeat_region	/rpt_family="ERV1"	16136..16218
repeat_region	/rpt_family="ERV1"	16682..16837
repeat_region	/rpt_family="ERV1"	16851..16876
repeat_region	/rpt_family="AT-rich"	16857..17182
repeat_region	/rpt_family="Alu"	17178..17665
repeat_region	/rpt_family="L1"	17166..17919
repeat_region	/rpt_family="ERV1"	

note="Intron-exon boundaries defined in relation to mRNA KIA00743, found in Genbank Accession Number AB018286." /codon\_start=1 /product="neurexin III" /protein\_id="AA013621.1" /db\_xref="GI:4240566" /translation="PSPDILITLSSKSSDDEDFEAGESTGEGELVLPDLPATPTATATPAPLTPPPFRFRLVLIETTRKSLNMTSAGVPCPLSSDGGDCDDGGIVTSYV

```

repeat_region      679.  .790      /rpt_family="TA)n"
repeat_region      complement(2076. 2106)
/rpt_family="LINE2"
repeat_region      2373.  .2397
/rpt_family="(TA)n"
repeat_region      complement(2410. 2671)
/rpt_family="AluCo"
unsure             2470.  .2480
/note="low quality data"
repeat_region      3439.  .3690
/rpt_family="AluSg1"
repeat_region      complement(3691. 3729)
/rpt_family="(CA)TA)n"
repeat_region      4215.  .4242
/rpt_family="AT_rich"
repeat_region      complement(4670. 4714)
/rpt_family="AT_rich"
repeat_region      complement(4727. 4872)
/rpt_family="AluUb"
repeat_region      complement(7014. 7315)
/rpt_family="AluY"
repeat_region      7545.  .7578
/rpt_family="AT_rich"
repeat_region      8414.  .8504
/rpt_family="(TA)n"
repeat_region      complement(8735. 8806)
/rpt_family="(TA)n"
repeat_region      10977.  .11055
/rpt_family="(GAA)n"
repeat_region      11084.  .11480
/rpt_family="MER63B"
repeat_region      complement(11482. 11655)
/rpt_family="MIR"
repeat_region      complement(12496. 12786)
/rpt_family="AluSx"
repeat_region      complement(12936. 13146)
/rpt_family="MIR"
repeat_region      complement(13488. 13786)
/rpt_family="AluY"
repeat_region      14532.  .14759
/rpt_family="MERS58A"
repeat_region      complement(14917. 15216)
/rpt_family="AluSg"
repeat_region      complement(15955. 16187)
/rpt_family="11PA14"
repeat_region      complement(17103. 17137)
/rpt_family="POL1A"
repeat_region      complement(17213. 17300)
/rpt_family="MIR"
repeat_region      complement(17325. 17360)
/rpt_family="AT_rich"
repeat_region      18242.  .18281
/rpt_family="(CAA)n"
repeat_region      18856.  .19028
/rpt_family="AluUb"
repeat_region      complement(20629. 20660)
/rpt_family="POLY A"
repeat_region      complement(21435. 2126)
/rpt_family="AluSx"
repeat_region      complement(21751. 22193)
/rpt_family="LINC4"
repeat_region      complement(22480. 22508)
/rpt_family="MIR"
repeat_region      complement(22495. 22632)
/rpt_family="LINE2"
repeat_region      complement(22904. 23125)
/rpt_family="MSTR"

```

```

repeat_region complement(23126..23429)
                /rpl_family="AT_rich"
repeat_region complement(23430..23595)
                /rpl_family="NSTA"
repeat_region complement(23772..24072)
                /rpl_family="ALUSx"
repeat_region complement(27422..27598)
                /rpl_family="MER2"
repeat_region complement(27906..28113)
                /rpl_family="MER20"
unsure complement(28100..28200)
                /note="low quality data"
repeat_region complement(28333..28353)
                /rpl_family="AT_rich"
repeat_region complement(30347..30515)
                /rpl_family="LIM6"
repeat_region complement(30499..30755)
                /rpl_family="LIM8"
repeat_region complement(34668..34742)
                /rpl_family="MER57_internal"
repeat_region complement(34901..35194)
                /rpl_family="L1"
repeat_region complement(35381..35448)
                /rpl_family="(CAT)n"
repeat_region complement(37534..37819)
                /rpl_family="LINE2"
repeat_region complement(38029..38595)
                /rpl_family="LIM9"
repeat_region complement(38601..39187)
                /rpl_family="TIGER1"
repeat_region complement(39198..39562)
                /rpl_family="LIP2"
repeat_region complement(39565..40986)
                /rpl_family="TIGER1"
repeat_region complement(40994..41121)
                /rpl_family="(GA)n"
repeat_region complement(41124..41217)
                /rpl_family="(GA)n"
repeat_region complement(41222..41361)
                /rpl_family="(GA)n"
repeat_region complement(41280..41450)
                /rpl_family="(GA)n"
unsure complement(41379..41710)
                /note="low quality data"
repeat_region complement(41379..41710)
                /rpl_family="TIGER1"
repeat_region complement(41717..41738)
                /rpl_family="AT_rich"
repeat_region complement(43779..43994)
                /rpl_family="MER30"
repeat_region complement(44206..44701)
                /rpl_family="LINE2"
repeat_region complement(44702..44948)
                /rpl_family="L1"
repeat_region complement(44801..45256)
                /rpl_family="LIM4"
repeat_region complement(46920..46948)
                /rpl_family="AT_rich"
unsure complement(47640..47740)
                /note="low quality data"
repeat_region complement(47669..47714)
                /rpl_family="(CA)n"
repeat_region complement(48325..48290)
                /rpl_family="(TA)n"
repeat_region complement(49741..49871)
                /rpl_family="(TAA)n"
repeat_region complement(50447..50501)
                /rpl_family="(GA)n"
repeat_region complement(51063..50983)
                /rpl_family="AT_rich"
repeat_region complement(52039..52067)
                /rpl_family="POLY_A"
repeat_region complement(52656..52955)
                /rpl_family="ALUY"
repeat_region complement(53075..53096)

```

```

repeat_region /rpl_family="AT_rich"
repeat_region /rpl_family="POLY_A"
repeat_region 58688..58779
repeat_region /rpl_family="(TAA)n"
repeat_region complement(58742..58783)
repeat_region /rpl_family="(TA)n"
repeat_region complement(60184..60483)
repeat_region /rpl_family="ALUY"
repeat_region complement(60629..60652)
repeat_region /rpl_family="AT_rich"
repeat_region 60904..60924
repeat_region /rpl_family="AT_rich"
repeat_region complement(61082..61360)
repeat_region /rpl_family="ALUSx"
repeat_region complement(62520..62627)

Query Match 16.3%; Score 39.2; DB 9; Length 180104;
Best Local Similarity 60.2%; Pred. No. 2;
Matches 65; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 98 GGATTTTATGATGTTGTTTAAAGGCTTGTGACCTGCGCTTAAATTAACAAAGACTAAT 205
Db 177891 GGGCTCTCTTTTCTTTCTTTACAGAAATGCTACCTTGTCTAAATTAATACAGAA 177832

QY 158 AATGATTATTGCGAGAGCGTGGCGCTTAATATTAATTAACAAAGACTAAT 205
Db 177831 AGCGAGAAATCCGAAAGAGTACGAAATATTTTAAATAAAGACCAAT 177784

RESULT 9
AC109963 166434 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-13SE1,*** SEQUENCING IN PROGRESS
DEFINITION *** 82 unordered pieces.
ACCESSION AC109963.3 GI:21738204
VERSION AC109963.3
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

REFERENCE
AUTHORS 1 (Pages 1 to 166434)
Wuzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dedelich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinn,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dumbin,K.J.,
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunarrane,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlek,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Huliy,S., Hume,C., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvath,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,Y., Liu,M., Loussegod,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokweno,S., Oquh,M., Okunodu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

```

Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
 Rivas, M., Rojas, A., Rojurokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,  
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Tellro, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.

-----  
 TITLE  
 JOURNAL  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 166434)  
 Worley, K.C.  
 Direct Submission  
 Submitted (09-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 166434)  
 Worley, K.C.  
 Direct Submission  
 Submitted (13-JUN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:18860231.

-----  
 COMMENT  
 -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Project Information  
 Center project name: GOLI  
 Center clone name: CH230-139E1  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap, version 0.990329  
 Consensus quality: 66236 bases at least Q40  
 Consensus quality: 73589 bases at least Q30  
 Consensus quality: 80243 bases at least Q20  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 82 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1  
 1081: contig of 1080 bp in length  
 1180: gap of unknown length  
 1181: contig of 1138 bp in length  
 2320: gap of unknown length  
 2420: contig of 1152 bp in length  
 3572: gap of unknown length  
 3672: gap of unknown length  
 4958: contig of 1287 bp in length  
 5058: gap of unknown length  
 4959: gap of unknown length  
 6429: contig of 1371 bp in length  
 6430: gap of unknown length  
 7725: contig of 1195 bp in length  
 7724: gap of unknown length  
 7825: contig of 1060 bp in length  
 8884: gap of unknown length  
 8885: gap of unknown length  
 8985: contig of 1271 bp in length  
 10255: gap of unknown length  
 10356: contig of 1302 bp in length  
 11658: gap of unknown length  
 11757: contig of 1122 bp in length  
 12979: gap of unknown length  
 12880: contig of 1123 bp in length  
 12980: contig of 1123 bp in length

14105: gap of unknown length  
 14205: contig of 1078 bp in length  
 15283: gap of unknown length  
 15383: contig of 1216 bp in length  
 16599: gap of unknown length  
 16699: contig of 1486 bp in length  
 18185: gap of unknown length  
 18285: contig of 1287 bp in length  
 19572: gap of unknown length  
 19672: contig of 1232 bp in length  
 20904: gap of unknown length  
 21004: contig of 1089 bp in length  
 22093: gap of unknown length  
 22193: contig of 1107 bp in length  
 23300: gap of unknown length  
 23400: contig of 1105 bp in length  
 24505: gap of unknown length  
 24605: contig of 1202 bp in length  
 25807: gap of unknown length  
 25907: contig of 1087 bp in length  
 26994: gap of unknown length  
 27094: contig of 1416 bp in length  
 28510: gap of unknown length  
 28610: contig of 1265 bp in length  
 28675: gap of unknown length  
 29975: contig of 1255 bp in length  
 31329: gap of unknown length  
 31330: contig of 1284 bp in length  
 32614: gap of unknown length  
 32714: contig of 1183 bp in length  
 33897: gap of unknown length  
 33997: contig of 1672 bp in length  
 35669: gap of unknown length  
 35769: contig of 1098 bp in length  
 36867: gap of unknown length  
 36967: contig of 1748 bp in length  
 38715: gap of unknown length  
 38815: contig of 1341 bp in length  
 40156: gap of unknown length  
 40256: contig of 1093 bp in length  
 41349: gap of unknown length  
 41449: contig of 1675 bp in length  
 43223: gap of unknown length  
 43124: contig of 1577 bp in length  
 43224: gap of unknown length  
 44800: contig of 1361 bp in length  
 44900: gap of unknown length  
 46797: contig of 1897 bp in length  
 46897: gap of unknown length  
 46898: contig of 1531 bp in length  
 48429: gap of unknown length  
 48529: contig of 1381 bp in length  
 50009: gap of unknown length  
 50010: contig of 1361 bp in length  
 51371: gap of unknown length  
 51470: contig of 1468 bp in length  
 52938: gap of unknown length  
 53038: contig of 1191 bp in length  
 54329: gap of unknown length  
 54330: contig of 1844 bp in length  
 56174: gap of unknown length  
 56274: contig of 1177 bp in length  
 57451: gap of unknown length  
 57550: contig of 1139 bp in length  
 58689: gap of unknown length  
 58789: contig of 1634 bp in length  
 60423: gap of unknown length  
 60424: contig of 1271 bp in length  
 61795: gap of unknown length  
 61894: contig of 1390 bp in length  
 63285: gap of unknown length  
 63384: contig of 1296 bp in length  
 64680: gap of unknown length  
 64780: contig of 1243 bp in length  
 64781: gap of unknown length  
 66023: contig of 1243 bp in length  
 66123: gap of unknown length

\* 66124 67877: contig of 1754 bp in length  
 \* 67878 67977: gap of unknown length  
 \* 67978 70618: contig of 2640 bp in length  
 \* 70618 70717: gap of unknown length  
 \* 70718 72859: contig of 2142 bp in length  
 \* 72860 72959: gap of unknown length  
 \* 72960 74745: contig of 1786 bp in length  
 \* 74746 74845: gap of unknown length  
 \* 74846 77563: contig of 2718 bp in length  
 \* 77564 79343: gap of unknown length  
 \* 79343 79350: contig of 1686 bp in length  
 \* 79350 81055: gap of unknown length  
 \* 79450 81055: contig of 1606 bp in length

Query Match 16.0%; Score 38.4; DB 2; Length 166434;

Best Local Similarity 56.2%; Pred. No. 3.4; Mismatches 56; Indels 0; Gaps 0;

Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 94 GGAGCGATTTTATGATGTTTGTAGAGCTTGTGACCTGCTTAAATACCTACAA 153

DB 11077 GGGGGGGGGTTTTTTTTTTTTTTTTTCCCTCCACCCCAAAAAA 11018

QY 154 ACAATATGATTTTTCGAGACCGTGGCGCTTATATAACAGAACTAATAACGCTC 213

DB 11017 AAAATTAATTTTTTTTTTGGGGGGGGGGAAGAAAAA 10958

QY 214 TGGCACC 221

DB 10957 CCCCCCCC 10950

RESULT 10

AC095187 192153 bp DNA linear HMG 11-JUN-2002

LOCUS AC095187.3 GI:21722694

DEFINITION Rattus norvegicus clone CH230-9C3, \*\*\* SEQUENCING IN PROGRESS \*\*\*

ACCESSION AC095187.3

VERSION HTG; HMG; PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 192153)

AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayalew, M., Banks, T.,

Barbata, J., Benton, J., Bimaga, K., Blankenburg, K., Bonini, D.,

Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,

Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Devilla, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,

Delaney, K.R., Delgado, O., Denna, A.L., Ding, Y., Dinh, H.H.,

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

Eanthart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Correll, J.H., Guerara, W., Gunaratne, P., Hale, S., Hamilton, K.,

Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,

Hernandez, O., Hodgson, A., Hogue, M., Hollaway, C., Hollins, B.,

Homs, F., Howard, S., Huber, J., Hui, J., Hui, J., Jackson, L.E.,

Jacobson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C.,

Kratovic, J., Kureishi, A., Landry, N., Leal, J., Lewis, L.C., Lewis, L.,

Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsbury, H.,

Lozano, R.J., Lu, X., Luster, A., Luster, R., Luster, R., Ma, J.,

Maneswar, M., Mapua, P., Martin, R., Martinale, A., Martinez, E.,

Masey, E., Maslany, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,

Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,

Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwona, G.,

Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Prims, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojokan, I., Rolfe, K., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shoshitaishvili, N., Sisson, I.,  
 Sledge, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Swales, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleceyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zortilla, S., Nelson, D.,  
 Weinstein, G., and Gbbs, R.

Unpublished

REFERENCE 2 (bases 1 to 192153)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (16-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 10, 2002 this sequence version replaced gi:17942286.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center clone name: CH230-9C3

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 119125 bases at least Q40

Consensus quality: 125091 bases at least Q30

Consensus quality: 130301 bases at least Q20

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)).

NOTE: This is a 'working draft' sequence. It currently

consists of 79 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 1243: contig of 1243 bp in length

1244 1243: gap of unknown length

1344 12721: contig of 1378 bp in length

2722 2821: gap of unknown length

2822 3941: contig of 1120 bp in length

3942 4041: gap of unknown length

4042 5289: contig of 1248 bp in length

5289 5389: gap of unknown length

5389 6956: contig of 1567 bp in length

6957 7056: gap of unknown length

7057 8133: contig of 1077 bp in length

8134 8234: gap of unknown length

8234 9265: contig of 1033 bp in length

9266 9365: gap of unknown length

9366 10944: contig of 1579 bp in length

10945 11044: gap of unknown length

11045 12750: contig of 1706 bp in length

12751 12851: gap of unknown length

12851 13958: contig of 1108 bp in length

13959 14058: gap of unknown length

```

* 14059 15275: contig of 1217 bp in length
* 15276 15375: gap of unknown length
* 15376 15793: contig of 1418 bp in length
* 16794 16893: gap of unknown length
* 16894 18019: contig of 1126 bp in length
* 18020 18119: gap of unknown length
* 18120 19768: contig of 1649 bp in length
* 19769 19868: gap of unknown length
* 19869 21147: contig of 1279 bp in length
* 21148 21247: gap of unknown length
* 21248 23000: contig of 1753 bp in length
* 23001 23100: gap of unknown length
* 23101 25073: contig of 1973 bp in length
* 25074 25173: gap of unknown length
* 25174 27094: contig of 1921 bp in length
* 27095 28757: gap of unknown length
* 27195 28857: contig of 1563 bp in length
* 28758 30242: gap of unknown length
* 28858 30342: contig of 1385 bp in length
* 30243 30343: gap of unknown length
* 30343 32113: contig of 1771 bp in length
* 32114 32213: gap of unknown length
* 32214 34175: contig of 1962 bp in length
* 34176 34275: gap of unknown length
* 34276 35471: contig of 1195 bp in length
* 35471 35570: gap of unknown length
* 35571 37322: contig of 1752 bp in length
* 37323 37422: gap of unknown length
* 37423 38791: contig of 1369 bp in length
* 38792 38891: gap of unknown length
* 38892 40361: contig of 1470 bp in length
* 40362 40461: gap of unknown length
* 40462 41995: contig of 1534 bp in length
* 41996 42095: gap of unknown length
* 42096 44157: contig of 2062 bp in length
* 44158 44257: gap of unknown length
* 44258 45517: contig of 1266 bp in length
* 45518 45617: gap of unknown length
* 45618 46970: contig of 1333 bp in length
* 46971 47070: gap of unknown length
* 47071 49593: contig of 2523 bp in length
* 49594 49693: gap of unknown length
* 49694 51492: contig of 1799 bp in length
* 51493 51592: gap of unknown length
* 51593 53377: contig of 1665 bp in length
* 53378 5378: gap of unknown length
* 5378 54782: contig of 1405 bp in length
* 54783 54882: gap of unknown length
* 54883 57021: contig of 2139 bp in length
* 57022 57121: gap of unknown length
* 57122 59001: contig of 1880 bp in length
* 59002 59101: gap of unknown length
* 59102 60121: contig of 1020 bp in length
* 60122 60221: gap of unknown length
* 60222 62559: contig of 2338 bp in length
* 62560 62659: gap of unknown length
* 62660 64766: contig of 2107 bp in length
* 64767 64866: gap of unknown length
* 64867 66866: contig of 1820 bp in length
* 66867 66786: gap of unknown length
* 66787 68406: contig of 1620 bp in length
* 68407 68506: gap of unknown length
* 68507 70230: contig of 1724 bp in length
* 70231 70330: gap of unknown length
* 70331 72687: contig of 2357 bp in length
* 72688 72787: gap of unknown length
* 72788 73961: contig of 1174 bp in length
* 73962 74061: gap of unknown length
* 74062 75594: contig of 1533 bp in length
* 75595 77376: gap of unknown length
* 75695 77376: contig of 1682 bp in length
* 77377 77476: gap of unknown length
* 77477 79269: contig of 1793 bp in length

```

```

* 79270 79370: gap of unknown length
* 79370 81510: contig of 2141 bp in length
* 81511 81510: gap of unknown length
* 81511 83780: contig of 2170 bp in length
* 83781 83781: gap of unknown length
* 83881 86085: contig of 2205 bp in length
* 86086 86185: gap of unknown length
* 86186 88050: contig of 1865 bp in length
* 88051 88151: gap of unknown length
* 88151 91051: contig of 2901 bp in length
* 91052 91151: gap of unknown length
* 91152 93335: contig of 2184 bp in length
* 93336 93435: gap of unknown length
* 93436 96781: contig of 3346 bp in length

Query Match
Best Local Similarity 55.1%; Score 38.4; DB 2; Length 192153;
Matches 75; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 101 TTTTAAAGATGTTTGTATAGAGCTTTGGTGACCTGCTAAATAATCAACAATAAT 160
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 31052 TTTTAAATACCTCTTATACCTGAATATCTACTGTAATAATCAACATGAATCT 30993
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

QY 161 GATTATTGCGAGACCGTGGCGCTTAATAAACAAGAACTAATAAGCGCTGCGACCC 220
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 30992 TATAGTACTGAGCTCTTGGTTTAATTTGAACAACAACAACAACAACAACA 30933
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

QY 221 CGCAATTAATATTTTGT 236
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
DB 30932 CACAAATTAACGTGT 30917
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

RESULT 11
AF022972/c 32940 bp DNA linear INV 17-MAY-2002
LOCUS
DEFINITION
ACCESSION AF022972
VERSION AF022972.1 GI:2384818
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 32940)
    Waterston R.
    Genome sequence of the nematode C. elegans: a platform for
    investigating biology. The C. elegans Sequencing Consortium
    Science 282 (5396), 2012-2018 (1998)
  2 (bases 1 to 32940)
    Graves, T., Wohlmann, P. and Gilling, B.
    The sequence of C. elegans cosmid F10D2
    Unpublished (2001)
  3 (bases 1 to 32940)
    Waterston, R.
    Direct Submission
    Submitted (17-MAY-2002) Department of Genetics, Washington
    University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
    Louis, MO 63110, USA
  Submitted by:
    Genome Sequencing Center
    Department of Genetics, Washington University
    St. Louis, MO 63110, USA, and
    Sanger Centre, Hinxton Hall
    Cambridge CB10 1RO, England
    email: rwenemacode.wustl.edu and jseesanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

```

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:  
<http://www.wormbase.org/db/seq/sequence?name=F10D2;class=Sequence>

#### NEIGHBORING COSMID INFORMATION

The 5' cosmid is F01C4, 997 bp overlap; the 3' cosmid is F09G2, 500 bp overlap. Actual start of this cosmid is at base position 798 of F10D2; actual end is at 6797 of F09G2.

#### NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFINDER (P. Green and L. Hillier, personal communication), the large scale EST projects of Yui Kobara ([http://www.ddb.jig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddb.jig.ac.jp/c-elegans/html/CE_INDEX.html)) and the C. elegans ORFome cloning project (<http://wormdb.dcl.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. TRNAS are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

#### FEATURES

```

source
    1. 32940
        /organism="Caenorhabditis elegans"
        /strain="Bristol N2"
        /db_xref="taxon:6239"
        /chromosome="V"
        /clone="F10D2"
        494..3628
            /gene="F10D2.3"
            /note="for a graphical representation of this gene see:
            http://www.wormbase.org/db/seq/sequence?name=F10D2.3;class=
            =Sequence"
            join(494..603,660..817,1066..1134,1233..1444,1558..2076,
            3092..3163,3235..3377,3424..3628)
            /gene="F10D2.3"
            /product="Hypothetical protein F10D2.3"
            /codon_start=1
            /protein_id="AAC48234.1"
            /translation="MLIHVNLTHRTSHNEFWDLKFLAECVLFIDLPNAIN
            FFLIRAHOFHNFECIYAMLEVINPTDIALMLKIMIMWEVDSPNTSMWLEFET
            MNTIACPAACMLFCFAVERCAVAYICDEKKORFETIMLSMGLIRFQAVCEY
            VTDRKNVYLLIFLIVDIGALMHCFLOWLNKRLYRLADNYITRYSILQRCVME
            NIYDGRKHILLISOKPHREFOMANNVITYMGPMNVYLLSVLFSDDILPHEHNYTL
            CLVSTIPIYRCXYPIYMHSCDWRKREINLPKPKIKIIPNYSKYPIRNSYGIS
            EENFMNHEHDHLDRAETNOSGRKPRYRIHRSKAKERSIEKFLKIRFHNAPSET
            AMSQMSPIRCCGIGCCPKTSKOSSTIRHSVFGMGCGFWPCIVLILSVETLN
            DVNHRITPWFSPSPRPATPRSTHEMOLNPSSTRHSRLHAY"
            complement(4273..6363)
            /gene="str-112"
            /note="for a graphical representation of this gene see:
            http://www.wormbase.org/db/seq/sequence?name=F10D2.4;class=
            =Sequence"
            complement(join(4273..4373,4427..4604,4712..4837,
            4883..4957,5004..5070,5123..5257,5322..5485,6169..6363))
            /gene="str-112"
            /codon_start=1
            /product="C. elegans STR-112 protein (corresponding
            sequence F10D2.4)"
            /protein_id="AAC48242.1"
            /db_xref="GI:2384829"

```

```

gene
    complement(7548..9883)
        /gene="F10D2.5"
        /note="for a graphical representation of this gene see:
        http://www.wormbase.org/db/seq/sequence?name=F10D2.5;class=
        =Sequence"
        complement(join(7548..7722,7838..7933,8029..8165,
        8211..8296,8419..8606,8652..8887,8954..9096,9143..9253,
        9387..9609,9655..9883))
        /gene="F10D2.5"
        /note="strong similarity to UDP-glucuronosyltransferases"
        /codon_start=1
        /product="Hypothetical protein F10D2.5"
        /protein_id="AAC48235.1"
        /db_xref="GI:2384822"
        /translation="MKGSLYLYLLISSTYSNIIVPCPLRGSHSPFGRADLL
        TEAGHNVLESPITIDERNYSYTKLTKQVAVIDPSELKALIGDLIAGKRWMOSES
        VEPIPTPREFKSITRQNNVANNALADGKOKKEDLILFSPVCCGLPAMELET
        KSPVAKOSIAYEPITIGADVYMSHIDPNIQAQSDKMLCGRMNFILQAMNVYG
        MPPEYQSYTPDEKTYTEIKLSAETFMNSPFLDPAPATTKNIOIGISYNDLP
        LKSGKLTPEMDQILNREKTLIVSRGYSITSDMPPRAYVGTITNMKQINDVPTIK
        YEDDKKEFLANKINIFHSKWVORLNLNNANFTEHTIVTESPRISAPWTHGL
        GYSDNEVSYFGKPTKCPISDQNRNKKMLERHNSIEFSYDHNKVAANFRLLY
        DESYTLASAKRLESEHLOFPYKPEELKIKHVEARFKPLSPIDPSRMQGFQYFLD
        IAYITVATVSYLFLVKFLFASCLKAKIKTD"
        complement(10003..13250)
        /gene="F10D2.6"
        /note="for a graphical representation of this gene see:
        http://www.wormbase.org/db/seq/sequence?name=F10D2.6;class=
        =Sequence"
        complement(join(10003..10192,10938..11033,11078..11214,
        11262..11347,11394..11533,11579..11951,12233..12363,
        12414..12723,13107..13250))
        /gene="F10D2.6"
        /note="strong similarity to UDP-glucuronosyltransferases;
        coded for by the following C. elegans cDNAs: YK1172610.3,
        YK1172610.5"
        /codon_start=1
        /product="Hypothetical protein F10D2.6"
        /protein_id="AAC48236.2"
        /db_xref="GI:20901939"
        /translation="MKLILFACFVNVNSYNFLFSPFGSHSTTFPAKIDLTSEA
        GHVTFEPPTIIRKRSKINYYKHKHVIHLEPSEKLERYGQMEDVDISRPWTDSSM
        AEMPMIKTFEMPRAEQAFVIGONLVYDEKEMKEDPMIPERRACAPALILEYIEIT
        TFIPSSILFEDNLISIGERIRFSTVYPKPKFDRKMTREBLAKIAIVPTFSLP
        THEFKSFRPYQALDIQMEHLSSFILSNSNYIDYPPYLTKNVQIGGISTVBNLK
        IHRVNMQRVIDMQRKTIIVSGSVYLSKDMVYNNKILAKTKMVFPPVITWKEF
        NDDULDEASTENIHSSVWQVQMLADRLRLFEFHAGLGVMVSYLQKPIQMPITP
        ADQMRNAKLANKNSIEISQYDNGEKELETLFTLFDISYRLSAEFLAQLQANP
        VKREKLLHAEFAQFGRPLSDPYSRQLSVEYFLIDVAAILIRPSVILLVYVFLQI
        IKLILNLFVTVDSRKYN"
        complement(13502..15416)
        /gene="F10D2.7"
        /note="for a graphical representation of this gene see:
        http://www.wormbase.org/db/seq/sequence?name=F10D2.7;class=
        =Sequence"
        complement(join(13602..13776,13825..14057,14101..14186,
        14250..14873,14917..15226,15273..15416))
        /gene="F10D2.7"
        /note="strong similarity to UDP-glucuronosyltransferases"
        /codon_start=1
        /product="Hypothetical protein F10D2.7"
        /protein_id="AAC48237.1"
        /db_xref="GI:2384824"
        /translation="MLIALFLFIYANVSSFWLFCPLRAHSHKFEAFITADSLTA
        GHVAFPPPIIVIEKRYRNPYTKSKYQVYVMESSKSKLGSQMSGTFDPRVYTESGA
        TEITIVYRIEKKMYNEQGVAKNIEIINDLAKRKEDVITLFEAFIRSAVPLMDYIEIP

```

```

TFPISLVADHDTALAVEIGEPVWSDCHSPDKLSTTERAFNAPWYFEDFMIG
YTHRSKPEPYNPIDISIPYPSAFVNSNPEDVPRPRLPTTGIGISIDNENR
SOKIDETWSEILIKREKTMVSGSMYSKDMPRKNAIAKNAKREKVMYFIRKED
DSDSEFAGCENIHGHSKIDUQSKRTIETARFATLHAGISITETSYCKRAICROD
EPOMKRTMAYRNGSITLSKIDUQSKRTIETARFATLHAGISITETSYCKRAICROD
IKPKEMWKAHEVAARFGRPLSDIPYSKRSMSVFETFLIDVAALCVETLTFVYVETL
VOSLFDILIKRKEK"
complement(16634, .18035)
/gene="F10D2.12"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=F10D2.12;class
s=Sequence"
complement(fjoin(16634, .16808,16852, .17084,17132, .17217,
17417, .18035))
/gene="F10D2.12"

CDS
Query Match 15.8% Score 38; DB 3; Length 32940;
Best Local Similarity 53.2%; Pred. No. 4.7;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 77 CATACTCCGAACACGCGATTTTATGATTTTGTGTTTAAAGCGCTTGGTCACCT 136
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9152 CACANACGAGGTGAGCTATATGATATTTTATGCTTAAATTTTGAAGCGT 9093

QY 137 GCCTAAATATCTACAAACAATATTTTGGAGAGCGCGCTTAAATATAACA 196
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9092 GATGGCACAACTCTGATTAATGAGCCTATGCGAAGAGATGTAACACTGATTC 9033

QY 197 AGAAGCTATAAAGC 210
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9032 AGCAATGCTCAATG 9019

RESULT 12
AC11374/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-81E15, *** SEQUENCING IN PROGRESS
AC11374
AC11374.2 GI:21735902
VERSION
HTG; HTGS; PHASE1.
KEYWORDS
Norway rat.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 114473)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooke,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Bileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavaros,S.R., Chacko,T., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dalhorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denu,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flaeg,N., Ford,J., Foster,P., Frantz,P.,
Gadisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,D., King,L., Korvah,J., Kovari,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Loulsegad,H.,
Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Maxhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

```

```

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Mckenzie,S., Ogih,M., Okunou,G.,
Oranque,N., Overedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scheer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Syatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tanev,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R., Wang,O.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 114473)
Worley,K.C.
REFERENCE
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 114473)
Worley,K.C.
REFERENCE
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18701138.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GMD
Center clone name: CH230-81E15
----- Summary Statistics
Sequencing Vector: Plasmid
Chemistry: Dye-terminator Big Dye 100k of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 69854 bases at least Q40
Consensus quality: 74696 bases at least Q30
Consensus quality: 78828 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/denbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1324: contig of 1324 bp in length
1325 1424: gap of unknown length
1425 2808: contig of 1484 bp in length
2809 3008: gap of unknown length
3009 4062: contig of 1054 bp in length
4063 4163: gap of unknown length
4163 5249: contig of 1087 bp in length
5250 5349: gap of unknown length
5350 6638: contig of 1289 bp in length
6639 6738: gap of unknown length
6739 7791: contig of 1053 bp in length
7792 7891: gap of unknown length
7892 8911: contig of 1020 bp in length
8912 9012: gap of unknown length
9012 10059: contig of 1048 bp in length
10060 10160: gap of unknown length
10160 11221: contig of 1062 bp in length
11221 11321: gap of unknown length

```

```

* 11322 12458: contig of 1137 bp in length
* 12459 12558: gap of unknown length
* 12559 13663: contig of 1105 bp in length
* 13664 13763: gap of unknown length
* 13764 14901: contig of 1138 bp in length
* 14902 15001: gap of unknown length
* 15002 16670: contig of 1669 bp in length
* 16671 16770: gap of unknown length
* 16771 18314: contig of 1544 bp in length
* 18315 18414: gap of unknown length
* 18415 20193: contig of 1779 bp in length
* 20194 20293: gap of unknown length
* 20294 21742: contig of 1449 bp in length
* 21743 21842: gap of unknown length
* 21843 23705: contig of 1863 bp in length
* 23706 23805: gap of unknown length
* 23806 25889: contig of 2084 bp in length
* 25890 27690: contig of 1701 bp in length
* 27691 27791: gap of unknown length
* 27792 29133: contig of 1343 bp in length
* 29134 29233: gap of unknown length
* 29234 30464: contig of 1231 bp in length
* 30465 30564: gap of unknown length
* 30565 32347: contig of 1783 bp in length
* 32348 32447: gap of unknown length
* 32448 33461: contig of 1014 bp in length
* 33462 34870: gap of unknown length
* 34871 34970: contig of 1309 bp in length
* 34971 36226: gap of unknown length
* 36227 36326: contig of 1256 bp in length
* 36327 37386: gap of unknown length
* 37387 37739: contig of 1412 bp in length
* 37739 37838: gap of unknown length
* 37839 39756: contig of 1918 bp in length
* 39757 39856: gap of unknown length
* 39857 41572: contig of 1716 bp in length
* 41573 41672: gap of unknown length
* 41673 43234: contig of 1562 bp in length
* 43235 43344: gap of unknown length
* 43345 44957: contig of 1623 bp in length
* 44958 45057: gap of unknown length
* 45058 46349: contig of 1292 bp in length
* 46350 46449: gap of unknown length
* 46450 48713: contig of 2264 bp in length
* 48714 48813: gap of unknown length
* 48814 49911: contig of 1099 bp in length
* 49912 50011: gap of unknown length
* 50012 52028: contig of 2017 bp in length
* 52029 52128: gap of unknown length
* 52129 54226: contig of 2098 bp in length
* 54227 54326: gap of unknown length
* 54327 56358: contig of 2033 bp in length
* 56359 56458: gap of unknown length
* 56459 57923: contig of 1465 bp in length
* 57924 58023: gap of unknown length
* 58024 60157: contig of 2134 bp in length
* 60158 60257: gap of unknown length
* 60258 61265: contig of 1008 bp in length
* 61266 61365: gap of unknown length
* 61366 62346: contig of 1181 bp in length
* 62347 62646: gap of unknown length
* 62647 64341: contig of 1695 bp in length
* 64342 64441: gap of unknown length
* 64442 66383: contig of 1942 bp in length
* 66384 66483: gap of unknown length
* 66484 67902: contig of 1419 bp in length
* 67903 68002: gap of unknown length
* 68003 69421: contig of 1419 bp in length
* 69422 69521: gap of unknown length
* 69522 71916: contig of 2395 bp in length
* 71917 72016: gap of unknown length
* 72017 74548: contig of 2532 bp in length

```

```

Query Match
Best Local Similarity 15.7%; Score 37.6; DB 2; Length 114473;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 102 TTTTATGATGTTTGTGTTTGAAGCGCTTGGTGAACCTGCTTAATAAATACATAAACAATG 161
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26389 TTTTTCGTTTGTGTTTGTGTTTGAAGTGGGTTAGCAATGTAAGACACACTGGAAGATC 26330
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 162 ATATATGCGAGACGCGGCGCTTAATATAACAGACACTAATAAAGCTGCGCAACC 221
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26329 ATTGTAGCAAAAGAAAGATGCGCTTATAGAAAGAAACAAAGCAACATACGGAAGT 26270
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 222 GCAA 225
      ||||
DB 26266 GCAA 26266

RESULT 13
AC007514/c 177720 bp DNA linear PRI 26-AUG-1999
DEFINITION Homo sapiens chromosome 14 clone BACs 209A20 and 396N8 map 14q31,
complete sequence.
ACCESSION AC007514
VERSION AC007514.5 GI:5776573
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 177720)
Dichhoff,R., Madan,A., Qin,S., Abbasi,N., Barradaran,L.,
Birchitt,B., Bloom,S., Dors,M., Rowen,L., Harrison,G., James,R.,
Lasky,S., Madan,A., Ratcliffe,A., Shaffer,T. and Hood,T.L.
Sequencing of human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177720)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dichhoff,R.,
James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S., Ratcliffe,A.,
Shaffer,T. and Hood,T.L.
Direct Submission
JOURNAL Submitted (08-MAY-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
AUTHORS 3 (bases 1 to 177720)
Dichhoff,R., Madan,A., Qin,S., Abbasi,N., Barradaran,L.,
Birchitt,B., Bloom,S., Dors,M., Rowen,L., Harrison,G., James,R.,
Lasky,S., Madan,A., Ratcliffe,A., Shaffer,T. and Hood,T.L.
Direct Submission
JOURNAL Submitted (26-AUG-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
On Aug 26, 1999 this sequence version replaced gi:568760.
This sequence is derived from the two haplotypes of RFL-11. No
genes were found in this sequence.
FEATURES
location/Qualifiers
1..177720
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```
source      /chromosome="14"
            /map="14q31"
            /clone="BACs 209A20 and 396N8"
            /clone_lib="RPC1-11"
            /note="These clones overlap BAC 526N18, Accession AF123462
            and BAC 115A24, Accession AC007372"
            1. 142843
source      /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /note="BAC 209A20"
            10153. 177720
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /note="BAC 396N8"
            1. 8135
misc_feature /note="Overlap with BAC 526N18, Accession AF123462"
            937
variation   /note="526N18: g; 209A20: t"
            /replace="t"
            1703
variation   /standard_name="t"
            /note="526N18: a; 209A20: t"
            1808
variation   /note="526N18: c; 209A20: t"
            /replace="t"
            3576
variation   /note="526N18: c; 209A20: t"
            /replace="t"
            7277
variation   /note="526N18: t; 209A20: c"
            /replace="c"
            12486
variation   /note="396N8: c; 209A20: a"
            /replace="a"
            13272
variation   /note="396N8: a; 209A20: t"
            /replace="t"
            13489
variation   /note="396N8: c; 209A20: t"
            /replace="t"
            13740
variation   /note="396N8: a; 209A20: g"
            /replace="g"
            16546
variation   /note="396N18: t; 209A20: c"
            /replace="c"
            17879
variation   /note="396N8: c; 209A20: t"
            /replace="t"
            18411
variation   /note="396N8: c; 209A20: t"
            /replace="t"
            19843
variation   /note="396N8: a; 209A20: g"
            /replace="g"
            19882
variation   /note="396N8: a; 209A20: g"
            /replace="g"
            20021
variation   /note="396N8: g; 209A20: t"
            /replace="t"
            20398
variation   /note="396N8: t; 209A20: c"
            /replace="c"
            20463
variation   /note="396N8: a; 209A20: g"
            /replace="g"
            21631
variation   /note="396N8: c; 209A20: t"
            /replace="t"
            21817
variation   /note="396N8: c; 209A20: t"
            /replace="t"
            42385

variation   21968
            /note="396N8: g; 209A20: t"
            /replace="t"
            22173
variation   /note="396N8: g; 209A20: a"
            /replace="a"
            23292
variation   /note="396N8: a; 209A20: t"
            /replace="t"
            24192
variation   /note="396N8: a; 209A20: t"
            /replace="t"
            25565
variation   /note="396N8: a; 209A20: g"
            /replace="g"
            26539
variation   /note="396N8: t; 209A20: c"
            /replace="c"
            27327
variation   /note="396N8: a; 209A20: g"
            /replace="g"
            28244
variation   /note="396N8: a; 209A20: g"
            /replace="g"
            28528
variation   /note="396N8: c; 209A20: t"
            /replace="t"
            28562
variation   /note="396N8: g; 209A20: a"
            /replace="a"
            28998
variation   /note="396N8: g; 209A20: a"
            /replace="a"
            29128
variation   /note="396N8: g; 209A20: a"
            /replace="a"
            29373
variation   /note="396N8: c; 209A20: a"
            /replace="a"
            29502
variation   /note="396N8: a; 209A20: t"
            /replace="t"
            31890
variation   /note="396N8: t; 209A20: c"
            /replace="c"
            34931
variation   /note="396N18: c; 209A20: t"
            /replace="t"
            37453
variation   /note="396N8: t; 209A20: c"
            /replace="c"
            37961
variation   /note="396N8: g; 209A20: a"
            /replace="a"
            38392
variation   /note="396N8: g; 209A20: a"
            /replace="a"
            38737
variation   /note="396N8: c; 209A20: t"
            /replace="t"
            39027
variation   /note="396N8: a; 209A20: g"
            /replace="g"
            40709
variation   /note="396N8: a; 209A20: t"
            /replace="t"
            41384
variation   /note="396N8: c; 209A20: a"
            /replace="a"
            41735
variation   /note="396N18: a; 209A20: acaca"
            /replace="acaca"
            42385
```



```

* 14992 16206: contig of 1215 bp in length
* 16207 16306: gap of unknown length
* 16307 17471: contig of 1165 bp in length
* 17472 17571: gap of unknown length
* 17572 18602: contig of 1031 bp in length
* 18603 18702: gap of unknown length
* 18703 20436: contig of 1734 bp in length
* 20437 20536: gap of unknown length
* 20537 21612: contig of 1076 bp in length
* 21613 21712: gap of unknown length
* 21713 23082: contig of 1380 bp in length
* 23083 23192: gap of unknown length
* 23193 24202: contig of 1010 bp in length
* 24203 25451: contig of 1149 bp in length
* 25452 25551: gap of unknown length
* 25552 26958: contig of 1407 bp in length
* 26959 27058: gap of unknown length
* 27059 29049: contig of 1991 bp in length
* 29050 29149: gap of unknown length
* 29150 30242: contig of 1093 bp in length
* 30243 30342: gap of unknown length
* 30343 31744: contig of 1402 bp in length
* 31745 31844: gap of unknown length
* 31845 32906: contig of 1062 bp in length
* 32907 33006: gap of unknown length
* 33007 34764: contig of 1758 bp in length
* 34765 34864: gap of unknown length
* 34865 36081: contig of 1217 bp in length
* 36082 36181: gap of unknown length
* 36182 37974: contig of 1793 bp in length
* 37975 38074: gap of unknown length
* 38075 39488: contig of 1414 bp in length
* 39489 39588: gap of unknown length
* 39589 40724: contig of 1136 bp in length
* 40725 40824: gap of unknown length
* 40825 42022: contig of 1198 bp in length
* 42023 42122: gap of unknown length
* 42123 44712: contig of 2590 bp in length
* 44713 44812: gap of unknown length
* 44813 46747: contig of 1935 bp in length
* 46748 46847: gap of unknown length
* 46848 48457: contig of 1610 bp in length
* 48458 48557: gap of unknown length
* 48558 50410: contig of 1853 bp in length
* 50411 50510: gap of unknown length
* 50511 51683: contig of 1173 bp in length
* 51684 51783: gap of unknown length
* 51784 53079: contig of 1296 bp in length
* 53080 53179: gap of unknown length
* 53180 54864: contig of 1685 bp in length
* 54865 54964: gap of unknown length
* 54965 56334: contig of 1370 bp in length
* 56335 56434: gap of unknown length
* 56435 58299: contig of 1865 bp in length
* 58300 58399: gap of unknown length
* 58400 60583: contig of 2184 bp in length
* 60584 60683: gap of unknown length
* 60684 63235: contig of 2552 bp in length
* 63236 63335: gap of unknown length
* 63336 64948: contig of 1613 bp in length
* 64949 65048: gap of unknown length
* 65049 66722: contig of 1674 bp in length
* 66723 66822: gap of unknown length
* 66823 68638: contig of 1816 bp in length
* 68639 68738: gap of unknown length
* 68739 70615: contig of 1877 bp in length
* 70616 70715: gap of unknown length
* 70716 72402: contig of 1687 bp in length
* 72403 72502: gap of unknown length
* 72503 74170: contig of 1668 bp in length
* 74171 74270: gap of unknown length
* 74271 77608: contig of 3338 bp in length

```

```

* 77609 77708: gap of unknown length
* 77709 80029: contig of 2321 bp in length
* 80030 80129: gap of unknown length
* 80130 82227: contig of 2098 bp in length
* 82228 82327: gap of unknown length
* 82328 84026: contig of 1699 bp in length
* 84027 84126: gap of unknown length
* 84127 87363: contig of 3237 bp in length
* 87364 89262: contig of 1799 bp in length
* 89263 89362: gap of unknown length
* 89363 90768: contig of 1406 bp in length
* 90769 90868: gap of unknown length
* 90869 92638: contig of 1770 bp in length

Query Match
Best Local Similarity 15.7%; Score 37.6; DB 2; Length 190481;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 102 TTTTATGATGTTTGTGTTTATGAGCGTTTGCGACCTGCTTAAATGAAATGCAAGCAAAATG 161
|||||
Db 178820 TTTTTCGTTTGTGTTTGTGTTTATGAGCGTTAGCAATGAAAGACACAGCAAGATC 178761
|||||

QY 162 ATATTTCGAGAGCGTGGCGCTTAAATGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 221
|||||
Db 178760 ATATTGAGCAAAAGAGATGCGTTAGTAAAGAAAACAAAGTCAAACTATACGGAAT 178701
|||||

QY 222 GCAA 225
Db 178700 GCAA 178697
|||||

RESULT 15
AC094816 220181 bp DNA linear HTG 08-JUL-2002
LOCUS Rattus norvegicus clone CH230-544, *** SEQUENCING IN PROGRESS ***,
AC094816 83 unordered pieces.
ACCESSION AC094816.4 GI:21702754
VERSION HTG: HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 220181)
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amarantunge,H.C., Aze,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,T., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dethorre,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,X., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Huylk,S., Hunne,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karissom,E., Kelly,S., Khan,V., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Locado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Na,D.,
Maheshwari,M., Mapua,P., Martin,R., Martinale,A., Martinez,E.,
Massey,E., Mashinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabdal,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,

```



```
* 69704 72322: contig of 2619 bp in length
* 72323 72422: gap of unknown length
* 72423 74890: contig of 2468 bp in length
* 74891 74990: gap of unknown length
* 74991 77937: contig of 2947 bp in length
* 77938 78037: gap of unknown length
* 78038 80271: contig of 2234 bp in length
* 80272 80371: gap of unknown length
* 80372 83452: contig of 3081 bp in length
* 83453 83552: gap of unknown length
* 83553 84865: contig of 1313 bp in length
* 84866 84965: gap of unknown length
* 84966 87255: contig of 2290 bp in length
* 87256 87355: gap of unknown length
* 87356 89706: contig of 2351 bp in length
```

Query Match 15.6%; Score 37.4; DB 2; Length 220181;  
Best Local Similarity 48.5%; Pred. NO. 6.1; Indels 0; Gaps 0;  
Matches 95; Conservative 0; Mismatches 101;

```
QY 26 AAAATCGTTGTACTAGACTTAGGCTGTTGGCAAGATCTTTAGCAACATATCCGAA 87
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1295 AAAAAATTTCCGAGGGGGGGGGGGGGGGGAAAAATTTTGGGGGGGGGNNNAAA 1354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 88 ACCACTGGACGAGATTTTANGATGTTGTTTAAAGGCTTGGTGAACCTGCTAAAAAFA 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1355 AAAACCCCCCCCCCCCCCTTTTCTTTTGTGGCCCCCCCCCCCCCACAACAAA 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 148 CTGCAACAAATGATTAATTGGAGAGCGGTGAATATATATAAACAAGAACTAATA 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1415 AAAAAAATAAATAAATTTGGGGGGGGGGGGGTTTAAAAAATAAATAAATAAATA 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 208 ACCTCTCGCAACCCGC 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1475 NNTTTTCCCCCCCCCCC 1490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 12, 2003, 01:46:42  
Job time : 1711 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compygen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:45:59 ; Search time 422.571 Seconds  
(without alignments)  
1279.025 Million cell updates/sec

Title: US-09-980-277-1  
Perfect score: 240  
Sequence: 1 gatcgaacgcacacacacg.....cgcaatattttttctt 240

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	44.2	18.4	120	AA170400
2	44.2	18.4	6194	Schistosoma mansoni
3	36.6	15.2	6134	GFP-Sm1-7 fusion p
4	34.2	14.3	581	Tumour suppressor
5	34	14.2	24	ABN60970
6	34	14.2	23	Human cancer relat
7	33.8	14.1	23	ABLI7579
8	33.6	14.0	16	Drosophila melanog
9	33.6	14.0	20	Haemolysin regulat
			20	B. burgdorferi ant
			20	B. burgdorferi ant

10	33.6	14.0	10723	22	AA14603	Dengue virus (DEN)
11	33.6	14.0	10723	22	AA14604	Dengue virus (DEN)
12	33.6	14.0	10735	22	AA14601	Wild type, virulen
13	33.6	14.0	10735	22	AA14602	Attenuated, vaccin
14	33.4	13.9	1860	24	ABN70879	Streptococcus poly
15	33.4	13.9	1950	24	ABN66701	Streptococcus poly
16	33.2	13.8	10949	22	AA107031	Human reproductive
17	33.2	13.8	10949	22	AAK79984	Human immune/haema
18	33.2	13.8	10949	22	AA162709	Human breast or ov
19	33.2	13.8	10951	22	AA107030	Human reproductive
20	33.2	13.8	10951	22	AAK79983	Human immune/haema
21	33.2	13.8	10951	22	AA162708	Human breast or ov
22	33	13.5	8923	22	AA13398	Drosophila melanog
23	32.4	13.5	8923	22	AA546340	Tumour suppressor
24	32.4	13.5	48551	24	AA520800	Clostridium diffic
25	32.2	13.4	964	24	ABO70115	Listeria monocytog
26	32.2	13.4	964	24	ABO71100	Listeria monocytog
27	32.2	13.4	2440	21	AAK59153	Human secreted pro
28	32.2	13.4	9359	20	AAK20262	Borrelia burgdorfe
29	32	13.3	83391	24	ABO67094	Human angiogenesis
30	31.8	13.2	580073	18	AA158840	Mycoplasma genital
31	31.6	13.2	621	23	ABV54682	Human prostate exp
32	31.6	13.2	669	24	ABN68671	Streptococcus poly
33	31.6	13.2	1933	22	AAK35566	Human immune/haema
34	31.6	13.2	2303	21	AAK18094	Lung cancer associ
35	31.6	13.2	10616	22	AAK85565	Human immune/haema
36	31.6	13.2	10616	22	AAK85567	Human immune/haema
37	31.6	13.2	40324	24	ABO67149	Human angiogenesis
38	31.2	13.0	581	24	ABQ33362	Oligonucleotide fo
39	31.2	13.0	581	24	ABQ33363	Oligonucleotide fo
40	31.2	13.0	1888	22	AAH2873	C albicans apoptos
41	31.2	13.0	5611	21	AA549548	5611 bp Candida al
42	31.2	13.0	12393	24	AB133263	Human immune syste
43	31.2	13.0	640681	24	AB192787	Buchnera sp. genom
44	30.8	12.8	1984	21	AAK76857	Human ORF2412
45	30.8	12.8	6931	22	ABAI7762	Human nervous syst

## ALIGNMENTS

RESULT 1	AA170400	standard; DNA; 120 BP.
ID	AA170400;	
AC	AA170400;	
XX		
DF	07-JAN-2002	(first entry)
XX		
DE	Schistosoma mansoni	repeater unit DNA.
XX		
KW	Schistosomiasis; infection; diagnosis; ds.	
XX		
OS	Schistosoma mansoni.	
XX		
PH	Key	Location/Qualifiers
FT	primer_bind	complement (1..19)
FT		/*tag= a
FT		/note= "primer of AA170401"
FT	primer_bind	91..110
FT		/*tag= b
FT		/note= "primer of AA170402"
XX		
XX		
PN	WO200175148-A1.	
XX		
PD	11-OCT-2001.	
XX		
PF	04-APR-2001; 2001MO-BR00035.	
XX		
PR	04-APR-2000; 2000BR-0001536.	
XX		
PA	(FLOC-) FLOCUTZ FUNDACAO CRUZ OSWALDO.	
XX		

PT Eukaryotic diploid multicellular parasite useful as universal grafts  
PT for in vivo delivery of beneficial gene products in humans and animals  
PT involves transformation with a transgene -

PS Claim 1; SEQ ID No 286; 27pp; English



PA	(PEKE ) PE CORP NY.
XX	
P1	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
XX	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
PS	Claim 1; SEQ ID NO 4210; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
CC	sequences (AB101840-AB116175) and the encoded proteins
CC	(A057737-AB17072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp://wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 1396 BP; 365 A; 346 C; 388 G; 297 T; 0 other;
Query Match	14.2%; Score 34; DB 23; Length 1396;
Best Local Similarity	57.3%; Pred. No. 2;
Matches	61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY	51 GCGGTGGGCAACTCTTTAGCAACATCCTCCGAACACCAGTGGAGATTTCATGAT 110  D2 212 GCGGTGGTCCGATATTCGCCCATATGTAATFAGCAACATTCAGGCGATTATAT 153
QY	111 GTTGTTTAGAGCGTTTGGAGACCTGCCTAAAATCTACAAATA 156   Db 152 GCTGTATTTGCGATTTGGCCGCGCACCAAGTACACACACA 107
RESULT	
ID	AB117578/c
XX	AB117578 standard; DNA; 7729 BP.
KC	AB117578;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 4207.
XX	
KM	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ds.
OS	Drosophila melanogaster.
XX	
XX	WO200171042-A2.
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001MO-US09231.
XX	
FR	23-MAR-2000; 2000US-191637P.
XX	
FR	11-JUL-2000; 2000US-061415O.
XX	
PA	(PEKE ) PE CORP NY.
XX	
P1	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
XX	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	

XX	PS	Claim 1; SEQ ID NO 4207; 21bp + Sequence Listing; English.
CC	CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB07840-AB16175) and the encoded proteins (AAB57737-AB572072).
CC	CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_ptc_sequences.
SQ	SQ	Sequence 7729 BP; 2361 A; 1594 C; 1587 G; 2187 T; 0 other;
Overy Match	14.2%; Score 34; DB 23; Length 7729;	
Best Local Similarity	57.5%; Pred. No. 3.2;	
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;		
Dy	51	GCTGTGGCAGATACCTTTTACCAACATACGCCGAAACCATGCAGCGATTATATGAT 110 
Db	1212	GCTGCTTTCCCGATATCCCGCATAGCTATTTGCAACAACATGCGCATTTATTTAT 1153 
Oy	111	GTTTGTTTAGAGCTTTGGTGACTGCTAAAATCTCAACAC 156 
Db	1152	GCTTGAATTCGCATTGGCCGCGGCGCACACAGTACACACAC 1107 
RESULT 7		
AAO75857 standard; DNA; 1020 BP.		
XX	XX	AAO75857;
DE	DE	06-JUL-1995 (first entry)
XX	XX	Haemolysin regulator DNA.
DE	DE	Haemolysin regulator; tuberculosis; vaccine;
KW	KW	antimense; therapy; ds.
XX	XX	Escherichia coli.
OS	OS	location/Qualifiers
PH	PH	Key
PH	PH	CDS
FT	FT	/tag= a
FT	FT	/note= "haemolysin regulator"
FT	FT	/tag= b
FT	FT	/note= "translation start codon"
FT	FT	-10_signal
FT	FT	/tag= c
FT	FT	/note= "5/6 match with E. coli promoter -10 consensus sequence (TATAAT)"
FT	FT	-35_signal
FT	FT	/tag= d
FT	FT	/note= "4/6 match with -35 consensus sequence (TTGACA)"
FT	FT	RBS
FT	FT	/tag= e
FT	FT	/note= "4/8 match with RBA binding site (AGGAAAGG)"
XX	XX	WO9428137-A.
PN	PN	08-DEC-1994.
PD	PD	25-MAY-1994; 94WO-GB05869.
PE	PE	24-MAY-1993; 93US-0066830.
XX	XX	(USSH ) US DEPT HEALTH & HUMAN SERVICE.

XX King CH, Sathish M, Shimnick TM;  
 XX  
 DR WPI: 1995-022809/03.  
 DR P-PSDB: AAR67645.  
 XX  
 PT DNA encoding cardapenem R gene product - used for production of  
 PT cardapenem antibiotics in bacteria and for bioassay.  
 XX  
 PS Claim 1; Page 38-39; 46pp; English.  
 XX  
 CC The sequence encodes the E. coli haemolysis protein regulator gene  
 CC which may be used, in an avirulent Mycobacterium bovis BCG vaccine  
 CC strain, to generate humoral and cellular responses against  
 CC Mycobacterium tuberculosis. Non-coding sequences may be coupled to  
 CC reporter genes to screen for compounds that increase gene  
 CC expression. DNA fragments can be used as probes for antisense  
 CC therapy to inhibit haemolysis, and the protein regulator gene  
 CC itself may be used as an immunogen for vaccines.  
 XX  
 SQ Sequence 1020 BP; 348 A; 195 C; 217 G; 260 T; 0 other;  
 XX  
 Query Match 14.1%; Score 33.8; DB 16; Length 1020;  
 Best Local Similarity 54.4%; Pred. No. 2.1;  
 Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
 QY 100 ATTTTATGATGTTGTTTAAAGCGCTTGAGTCCCTTAATAATCTCAACAA 159  
 DB 34 AATTTATTAATAGTTCTTAAACAGAGGTTTCATTACAATTTATATTAAAGGCGAA 93  
 QY 160 TGATTATTCGAGAGCGGCGGCTTAATAATAAAGAACTAATAACGCTCGCAC 219  
 DB 94 TGATTATGACTGAATCGTTGCAATTAACGGTAGAAGTATTAATAACCAATCGAAA 153  
 QY 220 CCGCA 224  
 DB 154 CCGCA 158  
 XX  
 RESULT 8  
 AAX61476  
 ID AAX61476 standard; DNA; 3363 BP.  
 XX  
 AC AAX61476;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein coding sequence, fl2.nt.  
 XX  
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.  
 XX  
 OS Borrelia burgdorferi.  
 OS  
 PN W09859071-AL.  
 PN  
 PD 30-DEC-1998.  
 PD  
 PF 18-JUN-1998; 98MO-US12718.  
 PF  
 PR 03-SEP-1997; 97US-0057483.  
 PR  
 PR 20-JUN-1997; 97US-0050359.  
 PR  
 PR 22-JUL-1997; 97US-0053344.  
 PR  
 PR 22-JUL-1997; 97US-0053377.  
 PR  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PA (MEDI-) MEDIMUNE INC.  
 PI Choi GH, Erwin AL, Hanson MS, Iathigra R;  
 XX  
 XX WPI: 1999-189980/16.  
 DR  
 DR P-PSDB: AAY19780.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop

PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by Borrelia, particularly Lyme disease  
 XX  
 PS Claim 1; Page 58-59; 275pp; English.  
 XX  
 CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.  
 XX  
 SQ Sequence 3363 BP; 1193 A; 388 C; 505 G; 1277 T; 0 other;  
 XX  
 Query Match 14.0%; Score 33.6; DB 20; Length 3363;  
 Best Local Similarity 50.0%; Pred. No. 3.4;  
 Matches 84; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
 QY 51 GCTGGTGGCAAGATCTTACGACATATCCGAACCACTGAGGATTTTATGAT 110  
 DB 259 GCAATATATCTTGAATAATTTCACTTATGACTTAAAGATGAAGTATATTATTTGAA 318  
 QY 111 GTTTGTTTAAAGCGCTTGAGACCTGCCATAAATACTAACAAAGATTAATTC 170  
 DB 319 GGAAGGATTAATATCTTGTGTAAGATATTAAAGAAATTAAGCAATATTAAAGGC 378  
 QY 171 GAGAGCGTGCGGCTTATATTAACAAGAACTAATAAAGCGCTCGCA 218  
 DB 379 GACAGAAATAGTCTTATTAAGAAGCTTAATAAAGAACTTATAGTATTTGA 426  
 XX  
 RESULT 9  
 AAX61475  
 ID AAX61475 standard; DNA; 3441 BP.  
 XX  
 AC AAX61475;  
 XX  
 DT 19-JUL-1999 (first entry)  
 XX  
 DE B. burgdorferi antigenic protein coding sequence, fl2.nt.  
 XX  
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.  
 XX  
 OS Borrelia burgdorferi.  
 OS  
 PN W09859071-AL.  
 PN  
 PD 30-DEC-1998.  
 PD  
 PF 18-JUN-1998; 98MO-US12718.  
 PF  
 PR 03-SEP-1997; 97US-0057483.  
 PR  
 PR 20-JUN-1997; 97US-0050359.  
 PR  
 PR 22-JUL-1997; 97US-0053344.  
 PR  
 PR 22-JUL-1997; 97US-0053377.  
 PR  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PA (MEDI-) MEDIMUNE INC.  
 PI Choi GH, Erwin AL, Hanson MS, Iathigra R;  
 XX  
 XX WPI: 1999-189980/16.  
 DR  
 DR P-PSDB: AAY19779.  
 XX  
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by Borrelia, particularly Lyme disease  
 XX  
 PS Claim 1; Page 57-58; 275pp; English.  
 XX  
 CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides

CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.

XX Sequence 3441 BP; 1219 A; 396 C; 513 G; 1313 T; 0 other;

Query Match 14.0%; Score 33.6; DB 20; Length 3441;  
 Best Local Similarity 50.0%; Pred. No. 3.4;  
 Matches 84; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 51 CTTGCTTGGCAGATACCTTTAGCAGACATCTCCGAAACCTGACGATTTTATGAT 110  
 DB 337 GCAGATATCTTGAAATTTTCATCTTACCTTACCAAGATTAATATTTTGA 396  
 QY 111 GTTGTTTAGAGCGCTTGACCTCCCTCAAAATACCAAAATGATATTTGC 170  
 DB 387 GGAAGATTATATCTGTGAGATATTAAGAAATTAAGACAAATTTAAAGSC 456  
 QY 171 GAGAGCGTGGCGTTATATATAAACAAGACTATATAACGCTCTCGCA 218  
 DB 457 GACGATATAGTCTCTATATAGACTCTAAAAAATTTATGCTATTGGA 504

# RESULT 10

AAD14603  
 ID AAD14603 standard; cDNA; 10723 BP.

XX AAD14603;

XX 01-NOV-2001 (first entry)

DE Dengue virus (DEN)-2/1-VP chimeric cDNA.

XX Flavivirus; Dengue virus; DEN; vaccine; infection; virucidal;

KW avirulent; immunogenic; viral disease; pharmaceutical; chimeric; ss.

XX Chimeric - Dengue virus type I.

XX Chimeric - Dengue virus type II.

XX Key Location/Qualifiers  
 FT CDS 97..10272  
 FT /tag= a  
 FT /product= "DEN-2/1-VP fusion protein"

PN W0200160847-A2.

XX 23-AUG-2001.

PD 16-FEB-2001; 2001WO-US05142.

PF 16-FEB-2000; 2000US-0182829.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA Kinney RM, Kinney CYH, Butrapet S, Gubler DL, Bhamarapravati N;

PI WPI; 2001-497162/54.

XX P-PSDB: AAE07982.

XX Chimeric flaviviruses that are avirulent and immunogenic, useful for

XX vaccinating against a range of dengue viruses -

XX Example 1; Page 155-170; 470pp; English.

CC The invention relates to avirulent, immunogenic flavivirus chimeras  
 CC comprising amino acid mutations in the non-structural proteins of a  
 CC flavivirus. Chimeric viruses containing the attenuated-mutated non-  
 CC structural genes of the virus are used as a backbone into which the  
 CC structural protein genes of a second flavivirus strain are inserted.  
 CC These chimeric viruses elicit pronounced immunogenicity but lack the  
 CC accompanying clinical symptoms of viral disease. Attenuated chimeric

CC flaviviruses are combined in a pharmaceutical composition to confer  
 CC simultaneous immunity against several strains of pathogenic flaviviruses  
 CC such as dengue virus serotypes DEN-1, DEN-2, DEN-3 and DEN-4. Immunogenic  
 CC flavivirus chimeras are also used as immunogens or multivalent vaccines  
 CC to confer simultaneous protection against infections. The present cDNA  
 CC sequence encodes dengue virus (DEN)-2/1-VP fusion protein related to the  
 CC invention. This fusion protein contains attenuated DEN-2 PDK-53V backbone  
 CC comprising a valine at the non-structural protein (NS3)-250 and the  
 CC capsid protein (C), premembrane/membrane protein (prM) and an envelope  
 CC protein (E) from wild-type DEN-1 16007 virus.

XX Sequence 10723 BP; 3514 A; 2229 C; 2722 G; 2258 T; 0 other;

Query Match 14.0%; Score 33.6; DB 22; Length 10723;  
 Best Local Similarity 50.0%; Pred. No. 4.6;  
 Matches 84; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 27 GAAATCGTTGATCTGACTTATGCTGCTGTTGGCAAGTCTTTTGCACATCTCCGA 86  
 DB 153 GAGAAACCGCGCTGTCACTCTTTCACAGTGGCGAGAGATTCGAAAAGATCTCTC 212  
 QY 87 AACCACTGAGACGAGATTTTATGATGTTGTTTAAAGCTTTGACCTGCTAAAT 146  
 DB 213 AGGCCAAGGACCAATGAAATGATGATGCTTCAATGCAATCTTAAGATTTCTAGCAT 272  
 QY 147 ACTACAAACAAATGATTTATTTCCGAGACGCTGGCGCTTAATATAA 194  
 DB 273 ACCCCCAACAGCAGGAATTTTGGCAGATGGGCTCATCAAGAGA 320

# RESULT 11

AAD14604  
 ID AAD14604 standard; cDNA; 10723 BP.

XX AAD14604;

XX 01-NOV-2001 (first entry)

DE Dengue virus (DEN)-2/1-VP chimeric cDNA.

XX Flavivirus; Dengue virus; DEN; vaccine; infection; virucidal;

KW avirulent; immunogenic; viral disease; pharmaceutical; chimeric; ss.

XX Chimeric - Dengue virus type I.

XX Chimeric - Dengue virus type II.

XX Key Location/Qualifiers  
 FT CDS 97..10272  
 FT /tag= a  
 FT /product= "DEN-2/1-VP fusion protein"

PN W0200160847-A2.

XX 23-AUG-2001.

PD 16-FEB-2001; 2001WO-US05142.

PF 16-FEB-2000; 2000US-0182829.

XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA Kinney RM, Kinney CYH, Butrapet S, Gubler DL, Bhamarapravati N;

PI WPI; 2001-497162/54.

XX P-PSDB: AAE07983.

XX Chimeric flaviviruses that are avirulent and immunogenic, useful for

XX vaccinating against a range of dengue viruses -

XX Example 1; Page 179-195; 470pp; English.

CC The invention relates to avirulent, immunogenic flavivirus chimeras  
 CC comprising amino acid mutations in the non-structural proteins of a



```

FT mutation replace (1608, C)
FT /tag- g
FT mutation replace (2363, A)
FT /tag- h
FT mutation replace (2695, T)
FT /tag- i
FT mutation replace (2782, C)
FT /tag- j
FT mutation replace (5063, G)
FT /tag- k
FT mutation replace (6048, A)
FT /tag- l
FT mutation replace (6806, A)
FT /tag- m
FT mutation replace (7330, A)
FT /tag- n
FT mutation replace (9445, C)
FT /tag- o
FT mutation /tag- o
WT0200160847-A2.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05142.
XX
XX 16-FEB-2000; 2000US-0182829.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Kinney RM, Kinney CYH, Butrapet S, Gubler DL, Bhamarapravati N;
XX P-PSDB; AAE07981.
XX
XX WPI: 2001-497162/54.
XX
XX Chimeric flaviviruses that are avirulent and immunogenic, useful for
XX vaccinating against a range of dengue viruses -
XX
XX Example 1; Page 130-146; 470pp; English.
XX
XX The invention relates to avirulent, immunogenic flavivirus chimeras
XX comprising amino acid mutations in the non-structural proteins of a
XX flavivirus. Chimeric viruses containing the attenuation-mutated non-
XX structural genes of the virus are used as a backbone into which the
XX structural protein genes of a second flavivirus strain are inserted.
XX These chimeric viruses elicit pronounced immunogenicity but lack the
XX accompanying clinical symptoms of viral disease. Attenuated chimeric
XX flaviviruses are combined in a pharmaceutical composition to confer
XX simultaneous immunity against several strains of pathogenic flaviviruses
XX such as dengue virus serotypes DEN-1, DEN-2, DEN-3 and DEN-4. Immunogenic
XX flavivirus chimeras are also used as immunogens or multivalent vaccines
XX to confer simultaneous protection against infections. The present CDNA
XX sequence encodes attenuated dengue-1 (DEN-1) PDK-13 virus protein variant
XX used for constructing flavivirus chimeras. Dengue virus types 1-4 (DEN-1
XX to DEN-4) are mosquito-borne flavivirus pathogens. The flavivirus genome
XX contains 5' non-coding region followed by a capsid protein (C) encoding
XX region, premembrane/membrane protein (pm) encoding region, an envelope
XX protein (E) encoding region followed by the region encoding non-
XX structural proteins (NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5) and finally a 3'
XX non-coding region.
XX
XX Sequence 10735 BP; 3429 A; 2231 C; 2776 G; 2299 T; 0 other;
XX
XX Query Match 14.0%; Score 33.6; DB 22; Length 10735;
XX Best Local Similarity 50.0%; Pred. No. 4.6;
XX Matches 84; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```

```

OY 147 ACTACAAACAAATGATTATTTCGAGACCGGCGCTAATATATAA 194
DB 271 ACCCCCAACAGCAGAAATTGGCTAGATGGGCGCTATCAAGAA 318
RESULT 14
ABN70879/c
ID ABN70879 standard; DNA; 1860 BP.
XX
XX ABN70879;
XX
XX 01-JUL-2002 (first entry)
XX
XX Streptococcus polynucleotide SEQ ID NO 9671.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX
XX 24-NOV-2000; 2000GB-0028727.
XX
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin B;
XX WPI: 2002-352536/38.
XX P-PSDB; ABP30248.
XX
XX New Streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX
XX Claim 7; Page 4093; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX
XX Sequence 1860 BP; 619 A; 296 C; 407 G; 538 T; 0 other;
XX
XX Query Match 13.9%; Score 33.4; DB 24; Length 1860;
XX Best Local Similarity 52.5%; Pred. No. 3.3;
XX Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
OY 35 TTGACTAGACTAGCGCTGGCGAAGATCTTGGCAACATCTCCGAACCACTG 94

```

```

Db 1099 TTGCTGCAATACGGGCTTTAGATGCCAAATTCCTTTTTCACATCTTCTTACCTTCTT 1040
QY 95 GACGGATTTTATGATGATGTTTGTAGAGGCTTTGTGACCTGCTTAAATACTACAAA 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1039 GTGGATTTTCTTAATAGGAATCGATGAAAGCTTCGCAAGACAGCATTTGATTTTAA 980
QY 155 CAAAATGATTTATTTGCGAG 173
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 979 CAACCTTCAGATTTCCAG 961
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
ABN6701/c
ID ABN6701 standard; DNA; 1950 BP.
AC ABN6701;
XX
XX
DE 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 1315.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KM group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KM antinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GE04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX
XX 24-NOV-2000; 2000GB-0028727.
XX
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin H;
XX
XX WPI: 2002-352536/38.
XX
XX P-PSDB; ABP26070.
XX
XX
XX New Streptococcus protein for the treatment or prevention of infection
XX or disease caused by Streptococcus bacteria, such as meningitis, and
XX for detecting a compound that binds to the protein -
XX
XX Claim 7; Page 3287-3288; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
XX the specification. The proteins have antibacterial and antinflammatory
XX activity. (I), nucleic acids encoding (II), ABN66044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX
XX Sequence 1950 BP; 656 A; 309 C; 428 G; 557 T; 0 other;

```

```

Best local Similarity 52.5%; Pred. No. 3.4;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 35 TTGATCTAGACTTAGCGTGTGGCAAGATCTTTAGCAACATACCGAAGCACTG 94
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1189 TTGCTGCAATACGGGCTTTAGATGCCAAATTCCTTTTCAAAATCTTCTTACCTTCTT 1130
QY 95 GACGGATTTTATGATGATGTTTGTAGAGGCTTTGTGACCTGCTTAAATACTACAAA 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1129 GTGGATTTTCTTAATAGGAATCGATGAAAGCTTCGCAAGACAGCATTTGATTTTAA 1070
QY 155 CAAAATGATTTATTTGCGAG 173
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1069 CAACCTTCAGATTTCCAG 1051
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Search completed: May 12, 2003, 01:14:51
Job time : 434.571 secs

```

Query Match

13.9%; Score 33.4; DB 24; Length 1950;

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - nucleic search, using SW model

Run on: May 12, 2003, 01:08:23 ; Search time 105.429 Seconds  
(without alignments)  
698.126 Million cell updates/sec

Title: US-09-980-277-1

Perfect score: 240  
Sequence: 1 gatctgaatcgacacacg.....cgcaattatattgtctt 240

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Database:

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PCrus.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	33.8	14.1	1020	1	US-08-557-115-2
2	33.8	14.1	1020	5	PCT-US94-05869-2
3	30.4	12.7	10718	3	US-08-325-428B-1
4	28.8	12.0	3047	1	US-08-280-690-1
5	28.6	11.9	1271	1	US-08-464-338A-1
6	28.6	11.9	1271	5	PCT-US94-14388-1
7	28.6	11.9	40138	3	US-09-090-793-12
8	28.6	11.9	162450	4	US-09-345-882-1
9	28.4	11.8	943	4	US-09-641-638-99
10	28.2	11.8	3785	4	US-09-889-718-1
11	28.2	11.8	7351	1	US-08-224-391-83
12	28.2	11.8	7351	1	US-08-484-304-83
13	28.2	11.8	7351	2	US-08-184-009-127
14	28.2	11.8	7351	2	US-08-566-398-39
15	28.2	11.8	7351	2	US-08-458-356-127
16	28.2	11.8	7351	2	US-08-658-665-39
17	28.2	11.8	7351	4	US-08-796-101-3
18	28.2	11.8	7351	4	US-08-460-736-127
19	28.2	11.8	7351	4	US-09-085-273-35
20	28.2	11.8	19011	1	US-08-310-356-36
21	28.2	11.8	19557	5	PCT-US92-06300-1
22	28	11.7	443	4	US-09-397-787-176
23	28	11.7	448	2	US-08-967-101-66
24	28	11.7	448	2	US-08-592-541-66
25	28	11.7	448	3	US-09-124-698-66
26	28	11.7	448	3	US-09-127-480-66
27	28	11.7	448	4	US-08-496-841C-66

c 28	28	11.7	448	4	US-09-124-523-66	Sequence 66, Appl
29	28	11.7	1049	2	US-08-961-527-317	Sequence 317, App
c 30	28	11.7	4184	2	US-08-785-310A-4	Sequence 4, Appl
31	27.8	11.6	1700	2	US-08-897-340-4	Sequence 4, Appl
32	27.8	11.6	1700	3	US-09-252-329-4	Sequence 4, Appl
33	27.6	11.5	1630	4	US-09-175-928-17	Sequence 17, Appl
c 34	27.6	11.5	3418	4	US-09-193-562D-29	Sequence 29, Appl
35	27.6	11.5	7425	4	US-09-453-702B-212	Sequence 212, App
c 36	27.6	11.5	11309	4	US-08-961-527-108	Sequence 108, App
37	27.4	11.4	921	4	US-09-134-001C-1841	Sequence 1841, Ap
38	27.4	11.4	5099	1	US-08-487-890A-4	Sequence 4, Appl
39	27.4	11.4	5099	2	US-08-478-435-4	Sequence 4, Appl
40	27.4	11.4	5099	2	US-08-337-483-4	Sequence 4, Appl
41	27.4	11.4	5099	2	US-08-478-573-4	Sequence 4, Appl
42	27.4	11.4	5099	3	US-08-474-671-4	Sequence 4, Appl
43	27.4	11.4	5099	3	US-08-483-577A-4	Sequence 4, Appl
44	27.4	11.4	5099	3	US-08-897-438-4	Sequence 4, Appl
45	27.4	11.4	5099	4	US-08-637-654-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1

US-08-557-115-2  
Sequence 2, Application US/0857115  
Patent No. 5731151

GENERAL INFORMATION:

APPLICANT: Kind, Harold C.

APPLICANT: Sathish, Mundayoor

TITLE OF INVENTION: Regulator of Contact-Mediated Hemolysis

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Jones & Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: U.S.A.

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,115

FILING DATE: 26-JUN-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Greene, Jamie L.

REGISTRATION NUMBER: 32,467

REFERENCE/DOCKET NUMBER: 03063-017105

TELEPHONE: (404) 818-3700

TELEFAX: (404) 818-3799

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1020 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORGANISM: E. coli

FEATURE:

NAME/KEY: misc.feature

LOCATION: 99..101

OTHER INFORMATION: /note="translation start codon"

FEATURE:

NAME/KEY: misc.feature

LOCATION: 99..101

OTHER INFORMATION: /note="translation start codon"

```

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: E. coli
FEATURE:
NAME/KEY: misc-feature
LOCATION: 99..101
OTHER INFORMATION: /note= "translation start codon"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 76..81
OTHER INFORMATION: /note= "5/6 match with -10
OTHER INFORMATION: consensus sequence for E. coli promoters (TATAAT)"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 47..52
OTHER INFORMATION: /note= "4/5 match with -35
OTHER INFORMATION: consensus sequence (TTGACA)"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 87..94
OTHER INFORMATION: /note= "4/8 match with the E. coli
OTHER INFORMATION: consensus sequence for ribosome binding sites
OTHER INFORMATION: (AGGAAAGG)"
FCT-US94-05869-2
Query Match 14.1%; Score 33.8; DB 5; length 1020;
Best Local Similarity 54.4%; Pred. No. 0.22;
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 100 ATTTTATGATGTTGGTTTGAAGAGCGTTGGTGACCTGCCCTAAATACTACAAACAAA 159
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 34 AATTATTAATAGTTGGTAAACAGAGAGTTTCATTACAAATTATTAATTAAGAGCGCA 93
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 160 TGATTATTTGCAGACGCGTGCGGCGTTAATATTAACAAAGACATAATAAGCGCTCGCGAC 219
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 94 TGATTATGACCTAATCGTTGCAGATTAACAGGTAGAAAGTGTATTAACAAAGCAATCGAA 153
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 220 CCGCA 224
| | | | |
DB 154 CCGCA 158
| | | | |
RESULT 3
US-08-325-4268-1
; Sequence 1, Application US/08325426B
; Patent No. 6017535
; GENERAL INFORMATION:
; APPLICANT: FO, Jianlin
; APPLICANT: TAN, Boon-Huan
; APPLICANT: YAP, Eu-Hian
; APPLICANT: CHAN, Yow-Chong
; APPLICANT: TAN, Yin-Hwee
; TITLE OF INVENTION: CDNA SEQUENCE OF DENGUE VIRUS SEROTYPE 1
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON AND VANDERHYTE PC
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,426B
; FILING DATE: 16-DEC-1994
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10718 base pairs

```

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA sequence corresponding to  
HYPOHETICAL: NO  
MOLECULE TYPE: the genomic RNA of DENV-S275/90  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Dengue Fever Virus Type 1  
STRAIN: S275/90  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 81..10268  
US-08-325-426B-1

Query Match 12.7%; Score 30.4; DB 3; Length 10718;  
Best Local Similarity 48.8%; Pred. No. 4.9;  
Matches 82; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Db 27 GAAATCGTGTATCTAGACTTAGGCTGTTGGCAAGATACCTTTAGCAACACTCCGA 86  
137 GAGAAACCGCGCTCACTGTTCAACATTTGGCGAAGAGATCTCAAAAGATTCCTTC 196  
QY 87 AACCACTGAGAGATTTTATGATTTTATAGAGCTTTGAGCTGACCTGGCTAAAT 146  
Db 197 AGCCAGAGACCCAGAAATTTGGATGCTTTCATAGCATCTCTAAGATTTCTAGCCAT 256  
QY 147 ACTACAAACAAATGATTTATTTGCGAGAGCGTGGCGCTTAATATATATAA 194  
Db 257 ACCCCACAGCAGGAATTTTGGCTAGATGGGCGCATTCAGAAAGA 304

RESULT 4  
US-08-280-690-1  
; Sequence 1, Application US/08280690  
; Patent No. 5583011  
; GENERAL INFORMATION:  
; APPLICANT: Mansour, Tag E  
; TITLE OF INVENTION: Compositions, Treatments, and  
; TITLE OF INVENTION: Diagnostics for Schistosomiasis and Related Diseases  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooler Godward  
; STREET: 5 Palo Alto  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94306-2155  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/280,690  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Neeley, Richard L  
; REGISTRATION NUMBER: 30,092  
; REFERENCE/DOCKET NUMBER: STAN-141/000US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-843-5070  
; TELEFAX: 415-857-0663  
; TELEX: 380816 COOLEYPA  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3047 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
; FEATURE:

NAME/KEY: CDS  
LOCATION: 147..2489  
US-08-280-690-1

Query Match 12.0%; Score 28.8; DB 1; Length 3047;  
Best Local Similarity 49.3%; Pred. No. 10;  
Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Db 51 GCTGTTGCAAGATCTTTAGCAACATCTCCGACAGAGATTTTATGAT 110  
Db 861 GATGATGGCGCTGAAATTTATGCTATTAAGTGAATGATGATGACATGACACAGT 920  
QY 111 GTTGTGTTAGAGGCTTTGGTGAAGCTGCTTAATAATCTCAACAAATGATTTTGC 170  
Db 921 GTCAATATATTAATTAAGTTGCTGCTAGAGTGTCTATCGATGACATGATGAACGATCCTGT 980  
QY 171 GAGAGCGTGGCGCTTAATATAAACAAGACT 202  
Db 981 GAATATGTAATAATTAATAGTATCAAGACT 1012

RESULT 5  
US-08-464-339A-1/c  
; Sequence 1, Application US/08464339A  
; Patent No. 5747280  
; GENERAL INFORMATION:  
; APPLICANT: HASTINGS, ET AL.  
; TITLE OF INVENTION: Human Vascular IBP-Like Growth  
; TITLE OF INVENTION: Factor  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/464,339A  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/14388  
; FILING DATE: 9 DEC 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MULILINS, J.G.  
; REGISTRATION NUMBER: 33,073  
; REFERENCE/DOCKET NUMBER: 325800-332  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1271 BASE PAIRS  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: CDNA  
US-08-464-339A-1

Query Match 11.9%; Score 28.6; DB 1; Length 1271;  
Best Local Similarity 48.7%; Pred. No. 9;  
Matches 73; Conservative 1; Mismatches 76; Indels 0; Gaps 0;

QY 13 ACCAAGCTTCTATGAAATCGTTGATCTAGACTTAGGCTGGTGGCAAGATCTTTTA 72  
Db 1240 ATCAGCAGCCTTTGTAATGCTCTCAGNTCAAAAGCTGTTGTACTCAAAATTCATTA 1181

```

Patent No. 6140486
GENERAL INFORMATION:
APPLICANT: Calgene, LLC
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression
of the gene for the enzyme acetyl-CoA carboxylase in plants
FILE REFERENCE: CGNE.131.0105
CURRENT APPLICATION NUMBER: US/09/090,793
PRIOR FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,650
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 40138
TYPE: DNA
ORGANISM: Vibrio marinus
US-09-090-793-12

Query Match      11.9%   Score 28.6;   DB 3;   Length 40138;
Best Local Similarity 50.4%;   Pred. No. 26;
Matches 70;   Conservative 0;   Mismatches 69;   Indels 0;   Gaps 0;

QY  96  ACGAGATTTTATGATGTTGTTTGGAGCGTTTGGTGAACCTGCTAAAATATCAAC 155
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  10727  ACGCTTTATATCAAGTGTGGAAGAAATTTGATTAACGTTGCACAAATGCGTGAT 10668

QY  156  AAAATGATTTATTCGAGAGCGTGGGCGCTTAATAATAAACAAGAACTAATAAAGCTGTC 215
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  10667  AAAAGCTTTGTAAGCAGAGTTTGGAGAGATGCACTTAACGTCAATTATGTGATGTA 10608

QY  216  GCACCCGCCAATTATATTTT 234
      |  ||  ||| ||| |||
Db  10607  CCTTACGGCGCTATATATAT 10589

RESULT 8
US-09-345-882-1
Sequence 1, Application US/09345882
Patent No. 639373
GENERAL INFORMATION:
APPLICANT: Bouquellet, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP)
FILE REFERENCE: GENSET 031A
CURRENT APPLICATION NUMBER: US/09/345,882
PRIOR FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 162450
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 72794
OTHER INFORMATION: 5-124-273 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 88073
OTHER INFORMATION: 5-127-261 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:

```

[illegible]

RESULT 11  
US-08-224-391-83/C  
; Sequence 83, Application US/08224391  
; Patent No. 5744140  
; GENERAL INFORMATION:  
; APPLICANT: Paoletti, Enzo  
; APPLICANT: Pincus, Steven E.  
; TITLE OF INVENTION: FLAVIVIRUS RECOMBINANT POXYTRUS VACCINE  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford  
; ADDRESSEE: c/o William S. Frommer  
; STREET: 530 Fifth Avenue  
;

CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/224,391  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/729,800  
FILING DATE: 17-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2340  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7351 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-224-391-83

Query Match 11.8%; Score 28.2; DB 1; Length 7351;  
Best Local Similarity 48.4%; Pred. No. 20;  
Matches 78; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 28 AAAATCGTGTATCTAGACCTTAGCGGTGGCAAGATACCTTTAGCAACATCTCCGAA 87  
DB 6654 ATAAAGTTAAATTTTATATACATACACAGGTGGAACAATCTTTCATGCTTAATTTCTTT 6595  
QY 88 ACCACTGACGAGATTTTATGATGTTGTTTGAAGGCTTGGTGAACCTCCCAAAAATA 147  
DB 6594 TTAAGTAGATATATTGTAAGATAATATTGAATTAATTAATTAATCAAAATATAGTC 6535  
QY 148 CTCACAACAAATGATTTTTCGCGAAGCGGTGGCGCTTAAT 188  
DB 6534 ATATAGTAGAGAGGATTTCCGTAGAGCTAGTAATATT 6494

RESULT 12  
US-08-484-304-83/c  
Sequence 83, Application US/08484304  
Patent No. 5744141  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: Pincus, Steven E.  
TITLE OF INVENTION: PLAYVIRUS RECOMBINANT POXYVIRUS VACCINE  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,304  
FILING DATE:

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/224,391  
FILING DATE:  
APPLICATION NUMBER: US 07/729,800  
FILING DATE: 17-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2340  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7351 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-484-304-83

Query Match 11.8%; Score 28.2; DB 1; Length 7351;  
Best Local Similarity 48.4%; Pred. No. 20;  
Matches 78; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 28 AAAATCGTGTATCTAGACCTTAGCGGTGGCAAGATACCTTTAGCAACATCTCCGAA 87  
DB 6654 ATAAAGTTAAATTTTATATACATACACAGGTGGAACAATCTTTCATGCTTAATTTCTTT 6595  
QY 88 ACCACTGACGAGATTTTATGATGTTGTTTGAAGGCTTGGTGAACCTCCCAAAAATA 147  
DB 6594 TTAAGTAGATATATTGTAAGATAATATTGAATTAATTAATTAATCAAAATATAGTC 6535  
QY 148 CTCACAACAAATGATTTTTCGCGAAGCGGTGGCGCTTAAT 188  
DB 6534 ATATAGTAGAGAGGATTTCCGTAGAGCTAGTAATATT 6494

RESULT 13  
US-08-184-009-127/c  
Sequence 127, Application US/08184009  
Patent No. 5833975  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: Tartaglia, James  
APPLICANT: Cox, William I.  
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY  
NUMBER OF SEQUENCES: 217  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/184,009  
FILING DATE: 19-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2330  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066CURFMS  
INFORMATION FOR SEQ ID NO: 127:

SEQUENCE CHARACTERISTICS:  
LENGTH: 7351 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-184-009-127

Query Match 11.8%; Score 28.2; DB 2; Length 7351;  
Best Local Similarity 48.4%; Pred. No. 20;  
Matches 78; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 28 AAAATCGTTGATCTAGACTTACGCTGGTGGCAGACTCTTTAGCAACATCTCCGAA 87  
DB 6654 ATAAAGTTTAAATTTTATATACACAGGTGGAACATCTTCATCGTATATCTTTT 6595  
QY 88 ACCACTGGACGAGATTTTATGATGTTGTTTACAGCCTTGGTACCGCTTAATAA 147  
DB 6594 TTAAGTAGATATATTTGTAGATATATTTGATTTACATATTTATATCAAAATTGTC 6535  
QY 148 CTACAAACAAATGATTTATTTGGAGAGCGTGGCGCTTAAT 188  
DB 6534 ATATAGTAAGAGAGATATTCGTAAGAGTCAGTAATAT 6494

RESULT 14  
US-08-566-398-39/c  
Sequence 39, Application US/08566398  
Patent No. 5858373  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIOUS  
TITLE OF INVENTION: PERITONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR  
NUMBER OF SEQUENCES: 63  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/566,398  
FILING DATE: 01-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2880  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-0712  
TELEFAX: (212) 840-3333  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7351 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-566-398-39

Query Match 11.8%; Score 28.2; DB 2; Length 7351;  
Best Local Similarity 48.4%; Pred. No. 20;  
Matches 78; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 28 AAAATCGTTGATCTAGACTTACGCTGGTGGCAGACTCTTTAGCAACATCTCCGAA 87

DB 6654 ATAAAGTTTAAATTTTATATACACAGGTGGAACATCTTCATCGTATATCTTTT 6595  
QY 88 ACCACTGGACGAGATTTTATGATGTTGTTTACAGCCTTGGTACCGCTTAATAA 147  
DB 6594 TTAAGTAGATATATTTGTAGATATATTTGATTTACATATTTATATCAAAATTGTC 6535  
QY 148 CTACAAACAAATGATTTATTTGGAGAGCGTGGCGCTTAAT 188  
DB 6534 ATATAGTAAGAGAGATATTCGTAAGAGTCAGTAATAT 6494

RESULT 15  
US-08-458-356-127/c  
Sequence 127, Application US/08458356  
Patent No. 594235  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: Tartaglia, James  
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY  
NUMBER OF SEQUENCES: 217  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,356  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/184,009  
FILING DATE: 19-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2530  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066CURTWS  
INFORMATION FOR SEQ ID NO: 127:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7351 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-458-356-127

Query Match 11.8%; Score 28.2; DB 2; Length 7351;  
Best Local Similarity 48.4%; Pred. No. 20;  
Matches 78; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 28 AAAATCGTTGATCTAGACTTACGCTGGTGGCAGACTCTTTAGCAACATCTCCGAA 87  
DB 6654 ATAAAGTTTAAATTTTATATACACAGGTGGAACATCTTCATCGTATATCTTTT 6595  
QY 88 ACCACTGGACGAGATTTTATGATGTTGTTTACAGCCTTGGTACCGCTTAATAA 147  
DB 6594 TTAAGTAGATATATTTGTAGATATATTTGATTTACATATTTATATCAAAATTGTC 6535  
QY 148 CTACAAACAAATGATTTATTTGGAGAGCGTGGCGCTTAAT 188  
DB 6534 ATATAGTAAGAGAGATATTCGTAAGAGTCAGTAATAT 6494

Tue May 13 09:58:32 2003

us-09-980-277-1.rni

Page 9

Search completed: May 12, 2003, 02:38:28  
Job time : 164.429 secs

---



; Sequence 9719, Application US/09764891  
; Publication No. US20030077808A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC006

; CURRENT APPLICATION NUMBER: US/09/764,891

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 10231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9719

; LENGTH: 10949

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-764-891-9719

Query Match 13.8%; Score 33.2; DB 9; Length 10949;  
Best Local Similarity 48.9%; Pred. No. 10;

Matches 89; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 11 CGACCAACCGCTTATGAAATCGTTGATCTAGACTTAGCGCTGGCAGATACCTT 70

DB 9927 CAACCTACTGATACATGTAACAAATGATGACCTCAGGGAATATGCTTGGGTGTT 9986

QY 71 TAGCAATATCTCCGAAACCACTGGAGCGATTTTATGATGTTGTTTGAAGCGCTTGG 130

DB 9987 TGGGGGGGACATCTTAAGTTTACACACGTATGATCTGCTTTTGAACATCTTAA 10046

QY 131 TGACCTGCTTAAATATCTACCAACAAATGATTTTTCGAGACGCTGGGCGTTATAT 190

DB 10047 TAATAAGCTATAGATGGGAACAGATTCGTGCTGCCAGCTCTTGGAGCTGAAGG 10106

QY 191 AA 192

DB 10107 AA 10108

### RESULT 3

US-09-764-891-9718

; Sequence 9718, Application US/09764891

; Publication No. US20030077808A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC006

; CURRENT APPLICATION NUMBER: US/09/764,891

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 10231

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9718

; LENGTH: 10951

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-764-891-9718

Query Match 13.8%; Score 33.2; DB 9; Length 10951;  
Best Local Similarity 48.9%; Pred. No. 10;

Matches 89; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 11 CGACCAACCGCTTATGAAATCGTTGATCTAGACTTAGCGCTGGCAGATACCTT 70

DB 9929 CAACCTACTGATACATGTAACAAATGATGACCTCAGGGAATATGCTTGGGTGTT 9988

QY 71 TAGCAATATCTCCGAAACCACTGGAGCGATTTTATGATGTTTGAAGCGCTTGG 130

DB 9989 TGGGGGGGACATCTTAAGTTTACACACGTATGATCTGCTTTTGAACATCTTAA 10048

QY 131 TGACCTGCTTAAATATCTACCAACAAATGATTTTTCGAGACGCTGGGCGTTATAT 190

DB 10049 TAATAAGCTATAGATGGGAACAGATTCGTGCTGCCAGCTCTTGGAGCTGAAGG 10108

QY 191 AA 192  
DB 10109 AA 10110

### RESULT 4

US-09-925-302-113/C

; Sequence 113, Application US/09925302

; Patent No. US20020044941A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA104

; CURRENT APPLICATION NUMBER: US/09/925,302

; CURRENT FILING DATE: 2001-08-10

; Prior application number: PCT/US00/05918

; Prior filing date: 2000-03-08

; Prior application number: 60/124,270

; Prior filing date: 1999-03-12

; NUMBER OF SEQ ID NOS: 896

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 113

; LENGTH: 2303

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-925-302-113

Query Match 13.2%; Score 31.6; DB 10; Length 2303;  
Best Local Similarity 53.2%; Pred. No. 15;

Matches 67; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 25 ATGAATATGCTGATCTAGACTTAGCTGCTGGCAAGATCTTTAGCAACATCTCC 84

DB 1105 ATGAATATGATTAATGGCTTATATCCCTGATTCATATTAATTAATAATTCCTA 1046

QY 85 GAACACCTGAGCGATTTTATGATGTTTGTAGAGCGCTTGTGACCTCCCTAAA 144

DB 1045 GAATATCTGCAAGATGATTAATTTAATCTTTAAGTAGCTGGAACCAACCAAAA 986

QY 145 ATACTA 150

DB 985 ATGTGA 980

### RESULT 5

US-09-815-923-9

; Sequence 9, Application US/09815923

; Publication No. US20020197644A1

; GENERAL INFORMATION:

; APPLICANT: Gill, Sarjeet S.

; APPLICANT: Ross, Linda S.

; TITLE OF INVENTION: Use of Insect Cell Membrane Transporters as No. US2002019764

; TITLE OF INVENTION: Target sites for insecticides

; FILE REFERENCE: 023070-03800US

; CURRENT APPLICATION NUMBER: US/09/815,923

; CURRENT FILING DATE: 2001-03-23

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 9

; LENGTH: 4956

; TYPE: DNA

; ORGANISM: Manduca sexta

; FEATURE:

; OTHER INFORMATION: neurotransmitter transporter encoded by inebriated

; OTHER INFORMATION: gene

; US-09-815-923-9

Query Match 13.1%; Score 31.4; DB 9; Length 4956;  
Best Local Similarity 54.9%; Pred. No. 25;

Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 48 TAGCTGCTGGCAAGATCTTTAGCAACATCTCCGAACCACTGGAGATTTTAT 107

Db 3488 TAGGCTGCTACACCTGTAACCTGTAAGTGTGTAACACCGAATTAATTTT 3547  
QY 108 GATGTTGTTTGAAGGCTTTGTGACCTGCTAAAAATACACAAATAAT 160  
3548 AATGTTGTGGAATTTGTGTGATCCCTACCGATAAATTTCAAAATACT 3600

RESULT 6  
US-09-938-842A-4228/c

; Sequence 4228, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRIPT300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 4228  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-4228

Query Match 13.0%; Score 31.2; DB 9; Length 2000;  
Best Local Similarity 63.2%; Pred. No. 19;  
Matches 48; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 97 CGAATTTTATGATGTTGTTTGAAGGCTTGGTGAACCTGCTAAATACTACAAACA 156  
Db 661 CAGACTATTTTGGAGCTGCTTTATATGCTGTTGAATTCGATATATATGATATATA 602  
QY 157 AATGATATTTGCGA 172  
Db 601 AATCATCATTCGCA 586

RESULT 7  
US-09-790-988-1

; Sequence 1, Application US/09790988  
; Patent No. US20020127687A1  
; GENERAL INFORMATION:  
; APPLICANT: SHIGENOBU, SHUJI  
; APPLICANT: WATANABE, HIDEMI  
; APPLICANT: HATORI, MASAHIRA  
; APPLICANT: SAKAI, YOSHIYUKI  
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
; FILE REFERENCE: 081356/0159  
; CURRENT APPLICATION NUMBER: US/09/790,988  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: JP2000-107160  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 640681  
; TYPE: DNA  
; ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match 13.0%; Score 31.2; DB 10; Length 640681;  
Best Local Similarity 48.8%; Pred. No. 2.6e+02;

Matches 84; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 35 TTGTACTAGACTTACGCTGGTGGCAAGATTACTTTAGCAACATACCGAATACACG 94  
Db 83625 TTTAAATGAAATATTAACCTTTAAAAAATCTTTTAAATTTTCCCTATATTTT 83684  
QY 95 GACGATTTTATGATGTTGTTTGAAGGCTTGTGACCTGCTAAAAATACATCAAA 154  
Db 83685 GTCTAAATATTAAGATTAATTTATTTTGGAGCCAGGTGATTTTAAAGTCATCACTAAAA 83744  
QY 155 CAAATGATTTTTCGAGACGCGGCGTTAATATAAAAACAAGATATTA 206  
Db 83745 AATTAAGATATTACTTCTACTGAGCTTTTAAATATAAAAAATATGAAA 83796

RESULT 8  
US-10-184-644-522/c

; Sequence 522, Application US/10184644  
; Publication No. US20030044930A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C227  
; CURRENT APPLICATION NUMBER: US/10/184,644  
; PRIOR FILING DATE: 2002-06-28  
; PRIOR APPLICATION removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 522  
; LENGTH: 527  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-184-644-522

Query Match 12.9%; Score 31; DB 9; Length 527;  
Best Local Similarity 10.3%; Pred. No. 12;  
Matches 21; Conservative 72; Mismatches 110; Indels 0; Gaps 0;

QY 38 TATCTAGACTTACGCTGGTGGCAAGATTACTTTAGCAACATACCGAATACACGAGAC 97  
Db 521 YAKM.M.GS...GMA...TATBGA...C.BH.SRD..WA.HDTTST.DMTCDYKX... 462  
QY 98 GATTTTATGATGTTGTTTGAAGGCTTGTGACCTGCTAAAAATACATCAAAACA 157  
Db 461 W.BYVH..NH..HDD.YS.YKTN.MRSSHA.BAY.TY..H.SAKAM.N..BTGCMWKOT 402  
QY 158 AATGATTTTTCGAGACGCGGGCGTTAATATAAAAACAAGATATTAAGCTCTCCG 217  
Db 401 NH..EC...BCBK.BCQR.T.R.CNRCCDA..TMAM.DC..HN..WHR.YANTC.AS. 342  
QY 218 ACCCGAATTAATATTTGTTCTT 240  
Db 341 MCMCRW.BM...T.TST..NT 319

RESULT 9  
US-10-184-634-522/c

; Sequence 522, Application US/10184634  
; Publication No. US20030068684A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc

```

; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Guiney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William L.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; PRIORITY FILING DATE: 2002-06-28
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-184-634-522

```

```

Query Match
Best Local Similarity 12.8%; Score 31; DB 9; Length 527;
Matches 21; Conservative 72; Mismatches 110; Indels 0; Gaps 0;

```

```

QY 38 TATCTAGACTTAGGCTGGTGGAGAGTACTTTTGGCAACATCTCGGAACACCTGGAC 97
DB 521 YAMN.M.GS...GMA...FATBGA...C.BH.SBD..WA.HPTSY.DMTCMDYKB... 462
QY 98 GGATTTTATGATGTTGTTTATAGAGCTTGGTGAACCTGCTGCTAAATAACACAAACAA 157
DB 461 W.BTYH..MB..HDD.YS.YTIN.MRSSH.BAY.TY..H.SAKAM.N.BTTCMTMDT 402
QY 158 AATGATTTTGGAGAGCGTGGCGTTAATATAACACAACTAATAACGCTTCGC 217
DB 401 NH..HC...BCBK.BCDR.T.R.CNKCD..TWAM.DC..HW...WHR.YANPC.AS. 342
QY 218 ACCCGAATTAATTTGTTCT 240
DB 341 MCMRYM.BM...T.TST..NT 319

```

```

RESULT 10
US-09-764-877-2137
; Sequence 2137, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2137
; LENGTH: 6931
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2137

```

```

Query Match
Best Local Similarity 12.8%; Score 30.8; DB 10; Length 6931;
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 20 GTTCATGAAATCGTTGATCTAGACTTAGGCTGGTGGCAACATCTTTAGCAACAT 79
DB 2843 GTGTTAGTAAAGATTAAATATGCACTGATGAGAGGTTGAAATCTTCATATAG 2902
QY 80 ACTCCGAACCACTGGAGAGATTTTATGATGTTTATGAGGCTTGTGACCTGCC 139
DB 2903 TGTGAGGACCTTAGCAACAATATATGAAATTTGCAACAAACGCTTCACAGCTTA 2962

```

```

QY 140 TAAAAATCTACAAACAAATGATTA 165
DB 2963 TAGTCATCTTCACACATATTGATA 2988

```

```

RESULT 11
US-09-764-877-2134
; Sequence 2134, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2134
; LENGTH: 15287
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2134

```

```

Query Match
Best Local Similarity 12.8%; Score 30.8; DB 10; Length 15287;
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```

```

QY 20 GTTCATGAAATCGTTGATCTAGACTTAGGCTGGTGGCAACATCTTTAGCAACAT 79
DB 1119 GTGTTAGTAAAGATTAAATATGCACTGATGAGAGGTTGAACTCTTCATATAG 11258
QY 80 ACTCCGAACCACTGGAGAGATTTTATGATGTTTATGAGGCTTGTGACCTGCC 139
DB 11259 TGTGAGGACCTTAGCAACAATATATGAAATTTGTCATCAAGCGTCTTCACAGCTTA 11318
QY 140 TAAAAATCTACAAACAAATGATTA 165
DB 11319 TAGTCATCTTCACACATATTGATA 11344

```

```

RESULT 12
US-09-790-988-1/C
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGEMORI, SHUJI
; APPLICANT: HATTORI, MASHIRA
; APPLICANT: WATANABE, HIDEMI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

```

```

Query Match
Best Local Similarity 12.8%; Score 30.8; DB 10; Length 640681;
Matches 65; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 46 CTTAGGCTGGTGGCAACATCTTTAGCAACATCTCGGAACCACTGGAGGATTTT 105
DB 609177 CTATCAAGATTTAAATATCATTTTAAATATTAATATTCACCTTAGGAGATCAT 609118
QY 106 ATGATGTTGTTTATGAGGCTTGGAGACCTGCTAAATAATACAAACAAATGATTA 165

```

Db 609117 TTGCACTGTGATTGGAATCGATTTAAATAATTTTCCATATAATTTTGAATAATGAATGTAA 609058

QY 166 TT 167  
 Db 609057 TT 609056

```

RESULT 13
US-09-998-598-363
Sequence 363. Application US/09998598
Patient No. US200201509222A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Medelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 363
LENGTH: 602
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-363

```

Query Match	12.5%;	Score 30;	DB 10;	Length 602;
Best Local Similarity	54.8%;	Pred. No. 25;		
Matches 57; Conservative	1;	Mismatches 46;	Indels 0;	Gaps 0;

QY 97 CGAGTTTAAAGATGTTTGTCTTTTAAAGAGCTTGTGTAACCTGCTAAATAAATCTACAAACA 156  
 Db 499 CGACACCTTAACCTTGTGTCTGTGTGKTKTTTGGACACCGTGTACAAAAA 558  
 QY 157 AATGATATTTCGAGAGCCGTGGGCGTATATTAACAAAGAA 200  
 Db 559 AGAAGATCTACTTGTAGCTTCGAGGATTTTAAAAA 602

```

? RESULT 14
? US-09-464-767-1/c
? Sequence 1, Application US/09464767
? Patent NO. US20020045249A1
?
? GENERAL INFORMATION:
?
? APPLICANT: Both, Gerald
? APPLICANT: Boyle, David
? APPLICANT: Vratil, Suchashnu
? TITLE OF INVENTION: DNA Encoding Ovine Adenovirus (OAV287) and Its Use as a Viral Vector
? FILE REFERENCE: 50179-073
?
? CURRENT APPLICATION NUMBER: US/09/464,767
? CURRENT FILING DATE: 1999-12-16
? NUMBER OF SEQ ID NOS: 3
?
? SOFTWARE: PatentIn version 3.0
?
? SEQ ID NO 1
?
? LENGTH: 29544
?
? TYPE: DNA
?
? ORGANISM: Ovine adenovirus
?
? US-09-464-767-1

```

Query Match	12.5%;	Score 30;	DB 10;	Length 29544;
Best Local Similarity	59.3%;	Pred. No. 1.5e+02;		
Matches	55;	Conservative	0;	Mismatches 35;
			Indels	0;
			Gaps	0;

QY	34	GTTGTA	CTCAG	ACTTAG	GCTGGT	GGCAAG	ATACTT	TTCGCA	ACAATCT	CCGAAC	CACT	93
Db	12259	GTTTAT	CTAC	CAATTA	ATCTGG	CGCAGA	AGCTAT	TACTGT	TAAAAA	ATTCGA	ACATTG	122000

94 GGACGATTTTATGATGTTGTTT 119

Db 12199 GGGCAATTGTTTTTATAAAAGTTT 12174

```

RESULT 15
US-09-464-767-3/C
; Sequence 3, Application US/09464767
; Patent No. US20020045249A1
; GENERAL INFORMATION:
; APPLICANT: Both, Gerald
; APPLICANT: Boyle, David
; APPLICANT: Viatli, Sudhansu
; TITLE OF INVENTION: DNA Encoding Ovine Adenovirus (OAV287) and Its Use as a Virus
; PRIORITY REFERENCE: 50179-073
; CURRENT APPLICATION NUMBER: US/09/464,767
; CURRENT FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 32745
; TYPE: DNA
; ORGANISM: synthetic construct
; US-09-464-767-3

```

Query Match	12.5%;	Score 30;	DB 10;	length 32745;
Best Local Similarity	59.3%;	Pred. No. 1.6e+02;		
Matches	51;	Conservative	0;	Mismatches 35;
			Indels	0;
			Gaps	0;

Accession	Sequence	Position
OY 34	GTGTGATCTAGACACTTAGGGCTGTGGCAAGATCTTTTACGACACATCTCCGGAACACT	93
Db 12258	GTTTCTTCTACCATTAATCTGGGTGAGAAAGCTATTACTGTTAAATAATCTGGAACACTG	12199
OY 94	GGACGGATTTTATGATGTTTGCTTTT	119
Db 12198	GGGCAATTTGTTTATATAAAGTTT	12173

Search completed: May 12, 2003, 04:47:58  
Job time : 530.571 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:04:28 ; Search time 2929.71 Seconds

(without alignments)  
1326.722 Million cell updates/sec

Title: US-09-980-277-1  
Perfect score: 240  
Sequence: 1 gatcgaatccgaccacacg.....cgcaattattttgtctt 240

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estha.\*  
2: em\_esthum.\*  
3: em\_estlin.\*  
4: em\_estlun.\*  
5: em\_estov.\*  
6: em\_estopl.\*  
7: em\_estro.\*  
8: em\_hlec.\*  
9: gb\_estl.\*  
10: gb\_estl2.\*  
11: gb\_hlc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estlun.\*  
16: em\_estlom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_huv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_frod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	44.2	18.4	687	17	BH210086 Sml-56J20
2	40.8	17.0	600	17	A0343444 RPI11-12
3	36.8	15.3	607	13	BI351574 IPL_72_F0
4	36.6	15.2	688	17	A2370732 IM0122E12
5	36.6	15.2	1099	17	CNS0016T
6	36.4	15.2	601	13	BML61288 EST563811

C 7	36.4	15.2	687	13	BH210086	687 bp	DNA	linear	GSS 24-OCT-2001
C 8	35.8	14.9	483	13	BH210086				
C 9	35.8	14.9	655	13	BH210086				
C 10	35.6	14.8	661	17	BH210086				
C 11	35.4	14.8	524	9	BH210086				
C 12	35.4	14.8	772	13	BH210086				
C 13	35.2	14.7	541	13	BH210086				
C 14	35.2	14.7	549	13	BH210086				
C 15	35.2	14.7	594	13	BH210086				
C 16	35.2	14.7	606	13	BH210086				
C 17	35.2	14.7	647	13	BH210086				
C 18	35.2	14.7	653	13	BH210086				
C 19	35.2	14.7	653	13	BH210086				
C 20	35.2	14.7	657	13	BH210086				
C 21	35.2	14.7	662	13	BH210086				
C 22	35.2	14.7	687	13	BH210086				
C 23	35.2	14.7	687	13	BH210086				
C 24	35.2	14.7	698	13	BH210086				
C 25	35.2	14.7	726	13	BH210086				
C 26	35.2	14.7	726	13	BH210086				
C 27	35.2	14.7	790	13	BH210086				
C 28	35.2	14.7	792	13	BH210086				
C 29	35.2	14.7	800	13	BH210086				
C 30	35.2	14.7	800	14	BH210086				
C 31	35.2	14.7	840	17	BH210086				
C 32	35.2	14.6	116	9	BH210086				
C 33	35.2	14.6	523	17	BH210086				
C 34	35.2	14.6	576	17	BH210086				
C 35	35.2	14.6	695	17	BH210086				
C 36	35.2	14.6	993	17	BH210086				
C 37	35.2	14.6	1101	17	BH210086				
C 38	34.8	14.5	474	17	BH210086				
C 39	34.8	14.5	609	13	BH210086				
C 40	34.8	14.5	1099	17	BH210086				
C 41	34.4	14.3	468	14	BH210086				
C 42	34.4	14.3	594	13	BH210086				
C 43	34.4	14.3	607	13	BH210086				
C 44	34.4	14.3	742	13	BH210086				
C 45	34.4	14.3	753	13	BH210086				

## ALIGNMENTS

RESULT 1  
LOCUS BH210086  
DEFINITION Sml-56J20.TF Sml Schistosoma mansoni genomic clone Sml-56J20, DNA sequence.  
ACCESSION BH210086.1 GI:16389299  
VERSION BH210086  
KEYWORDS GSS.  
SOURCE Schistosoma mansoni.  
ORGANISM Schistosoma mansoni.  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE 1 (bases 1 to 687)  
Shetty,V., Simpson,A., Malek,J., Koo,H., Loyverde,P.T. and El-Sayed ,N.M.  
Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sml BAC library for gene discovery and map construction  
Unpublished (2001)  
Other GSS: Sml-56J20.TF  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@igr.org  
10.edu)  
Seq primer: M13 For  
Class: BAC ends.

FEATURES  
SOURCE  
Location/Qualifiers  
1. 687  
/organism="Schistosoma mansoni"  
/strain="Puerto Rico"  
/db\_xref="taxon:6183"  
/clone="Sm1-56j20"  
/clone\_1db="Sm1"  
/note="Vector: pBelobAC11, Site\_1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelobAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 X the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

BASE COUNT  
218 a 101 c 152 g 216 t

ORIGIN

Query Match  
Best Local Similarity 18.4%; Score 44.2; DB 17; Length 687;  
Matches 46; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 81 CTCGAAACCACTGAGCGATTTTATGATGTTGTTTATGAGCGCTTGG 129  
|||||  
Db 388 CTCGAAACCACTGAGCGATTTTATGATGTTGTTTATGATTTATTTG 436

RESULT 2  
LOCUS  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC0343444  
AC0343444  
AC0343444.1 GI:4168340  
GSS.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 600)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,  
J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Unpublished (1997)  
Other\_GSS: RPCI11-124D13.T5  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetig@ig.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieder@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page:  
[http://www.tigr.org/tdb/human/gen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/human/gen/bac_end_search/bac_end_search.html)  
Seq primer: 17  
Class: BAC ends.  
Location/Qualifiers  
1. 600  
/organism="Homo sapiens"  
/db\_xref="GDB:7547316"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-124D13"  
/clone\_1db="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"

FEATURES  
SOURCE  
Location/Qualifiers  
1. 607  
/organism="Sorghum bicolor"  
/cultivar="Brx623"  
/db\_xref="taxon:4558"  
/clone\_1db="Immature panicle 1 (IP1)"  
/note="Organ: Developing preanthesis panicles; Vector:  
pBluescript II SK(-) from lambda Zap II; Site\_1: XhoI;  
Site\_2: EcoRI; The library was made from poly-A RNA in the  
cloning vector lambda Zap II. Clones to be sequenced were  
prepared by mass excision."

BASE COUNT  
142 a 159 c 142 g 164 t

ORIGIN

Query Match  
Best Local Similarity 17.0%; Score 40.8; DB 17; Length 600;  
Matches 109; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 22 TCTATGAAATGCTTATCTAGCTTAGCGGTGGCAAGATCTTTAGCAATAC 81  
|||||  
Db 45 TATATACCTCTTATAGATGATTTATATATAGAAATGTAATCATTTCAATTA 104

QY 82 TCCGAAACACGCGAGCGATTTTATGATGTTTGTATAGAGCTTGGTACCTGCTTA 141  
|||||  
Db 105 TCTATATGATGCTTATCTTATTAATTAATTTATTAAGCTTATGCACTTAAA 164

QY 142 AAATCTCAACAAATGATTTTGGAGAGCGTGGCGTTATATTAACAGAAC 201  
|||||  
Db 165 TGTATCTGAGAAATTAAGGTCTATGATAGTATATCACTTAATTAAGAGTAA 224

QY 202 TATTAACCGCTCGGACCGCAATTAATTT 233  
|||||  
Db 225 AGATTAACCTAACCCCTTACACACTAATAT 256

RESULT 3  
LOCUS  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BI351574/c  
BI351574  
BI351574.1 GI:15046016  
EST.  
Sorghum.  
Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC  
Clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 607)  
Klein,R.R., Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt  
L.H.  
An EST database from Sorghum: developing preanthesis panicles  
Unpublished (2001)  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Km. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmp@prattuga.edu  
Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for high quality sequence is  
20. Three-prime sequences, which are obtained with polyTmix or 17  
sequencing primer, are presented as the reverse complement.  
Seq primer: JBN REV  
High quality sequence stop: 599  
POLYA-No.  
Location/Qualifiers  
1. 607  
/organism="Sorghum bicolor"  
/cultivar="Brx623"  
/db\_xref="taxon:4558"  
/clone\_1db="Immature panicle 1 (IP1)"  
/note="Organ: Developing preanthesis panicles; Vector:  
pBluescript II SK(-) from lambda Zap II; Site\_1: XhoI;  
Site\_2: EcoRI; The library was made from poly-A RNA in the  
cloning vector lambda Zap II. Clones to be sequenced were  
prepared by mass excision."

BASE COUNT  
142 a 159 c 142 g 164 t

ORIGIN

Query Match  
Best Local Similarity 15.3%; Score 36.8; DB 13; Length 607;







RESULT 11  
 AL728080/c  
 LOCUS  
 DEFINITION AL728080 Dario rerio embryonic inner ear subtracted cDNA Dario  
 accession AL728080  
 version AL728080.1 GI:20192684  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM zebrafish.  
 Dario rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 ; Cyprinidae; Danio.  
 1 (bases 1 to 524)  
 Colubra, R., Weil, D., Brothier, P., Blanchard, S., Levi, M., Hardegin  
 , J.P., Weissendbach, J., and Petit, C.  
 A subtracted cDNA library from the zebrafish (Dario rerio)  
 embryonic inner ear  
 Unpublished (2002)  
 CONTACT: Genoscope  
 Genoscope, Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.  
 Location/Qualifiers  
 1..524  
 /organism="Dario rerio"  
 /db\_xref="taxon:7955"  
 /clone="BN08A0942B11"  
 /clone\_lib="Dario rerio embryonic inner ear subtracted  
 cDNA"  
 /tissue\_type="inner ear"  
 /dev\_stage="embryonic"  
 /note="subtracted cDNA library"  
 BASE COUNT 194 a 68 c 78 g 184 t  
 ORIGIN  
 Query Match 14.8%; Score 35.4; DB 9; Length 524;  
 Best Local Similarity 53.2%; Pred. No. 65;  
 Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
 QY 100 ATTTTATGATGTTGTTTGAAGGCTTGTGACCTGCTTAAATTAATCTACAAACAAA 159  
 DB 318 ATTTTATGATGTTGTTTGAAGGCTTGTGACCTGCTTAAATTAATCTACAAACAAA 159  
 QY 160 TGAATTTGCGAGAGCGGTGCGCTTAATTAACAGACAACTAATTAACGCTCGCAC 219  
 DB 258 GCTATGAAAAAATGAAAGTCATTTTAAACATTAATAAAAAAATTAACGTTTGAAC 199  
 QY 220 CCGCAATTAATTTGTTCTT 240  
 DB 198 ACTAAATTAATTAATTTT 178  
 RESULT 12  
 BU36150/c  
 LOCUS  
 DEFINITION BU36150 Dictyostelium discoidium cDNA library, AF Dictyostelium  
 accession BU36150  
 version BU36150.1 GI:19255745  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Dictyostelium discoidium.  
 Dictyostelium discoidium  
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 1 (bases 1 to 772)  
 Urushihara, H., Tanaka, Y., Kohara, Y., and Shin-i, T.  
 Full length cDNA of Dictyostelium discoidium at the aggregation  
 stage  
 Unpublished (2002)  
 CONTACT: Tadasi Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.  
 Location/Qualifiers  
 1..772  
 /organism="Dictyostelium discoidium"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone="da57j09"  
 /clone\_lib="Dictyostelium discoidium cDNA library, AF"  
 /sex="mat A"  
 /dev\_stage="Aggregation stage"  
 BASE COUNT 265 a 125 c 158 g 224 t  
 ORIGIN  
 Query Match 14.8%; Score 35.4; DB 13; Length 772;  
 Best Local Similarity 53.2%; Pred. No. 60;  
 Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
 QY 98 GGAATTTTGAATGTTGTTTGAAGGCTTGTGACCTGCTTAAATTAATCTACAAACAA 157  
 DB 153 GGCCTTACTGCTGATTTAGCTCAACCTATGGGATGATTTAGAAACCTTTAAACAA 94  
 QY 158 AATGATTATTTGCGAGAGCGGTGCGCTTAATTAACAGAACTAATTAACGCTCGC 217  
 DB 93 CAATTAATTAAGAAATTTGTAACAGATGATTCATAAAGATATACATATGCTCAT 34  
 QY 218 ACCCGCAATTAATTTGCTC 238  
 DB 33 ATGGCATTTTAATTAATTTTC 13  
 RESULT 13  
 BU444759/c  
 LOCUS  
 DEFINITION BU444759 Dictyostelium discoidium cDNA library, VF Dictyostelium  
 accession BU444759  
 version BU444759.1 GI:19419480  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM Dictyostelium discoidium.  
 Dictyostelium discoidium  
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 1 (bases 1 to 541)  
 Urushihara, H., Tanaka, Y., Kohara, Y., and Shin-i, T.  
 Full length cDNA of Dictyostelium discoidium at the vegetative  
 stage  
 Unpublished (2002)  
 CONTACT: Tadasi Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.  
 Location/Qualifiers  
 1..541  
 /organism="Dictyostelium discoidium"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone="da57j09"  
 /clone\_lib="Dictyostelium discoidium cDNA library, VF"  
 /sex="mat A"  
 /dev\_stage="Growth phase"  
 BASE COUNT 163 a 85 c 81 g 212 t  
 ORIGIN  
 Query Match 14.7%; Score 35.2; DB 13; Length 541;  
 Best Local Similarity 57.1%; Pred. No. 73;  
 Matches 64; Conservative 0; Mismatches 48; Indels 0; Gaps 0;  
 QY 125 CTTTGGTACCTGCTTAAATTAATCTACAAACAAATGATTTGCGAGAGCGTGGCGT 184



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:02:58 ; Search time 99 Seconds  
(without alignments)  
5879.355 Million cell updates/sec

Title: US-09-980-277-2

Perfect score: 20

Sequence: 1 gatcgaatcgaacacacg 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_com:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vt:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mus:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vt:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_man:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	121	3	SCMRPTAA
2	16.8	84.0	4676	3	DROMTBA
3	16.8	84.0	22663	3	DROMHC
4	16.8	84.0	68879	3	AC005119
5	16.8	84.0	79309	2	AC0019767
6	16.8	84.0	175413	3	AC093097
7	16.8	84.0	258166	3	AE003652
8	16.4	82.0	99542	2	AC095629
9	16.4	82.0	132927	8	AC079890
10	16.4	82.0	145668	9	AC092272
11	16.4	82.0	155613	9	AC019350
12	16.4	82.0	165237	2	AC091914
13	16.4	82.0	177246	2	AC114308
14	16.4	80.0	9359	2	AC018248
15	16.4	80.0	172069	3	AC012160
16	16.4	80.0	177793	3	AC012096
17	16.4	80.0	209280	2	AC116528
18	16.4	80.0	298514	3	AE003504
19	15.8	79.0	355	3	AE449752
20	15.8	79.0	754	9	HS4328235
21	15.8	79.0	1655	5	XLGBETAL
22	15.8	79.0	3671	6	AR171292
23	15.8	79.0	3671	6	AR172793
24	15.8	79.0	10019	1	AE008490
25	15.8	79.0	24519	2	AC020211
26	15.8	79.0	61009	2	AC084776
27	15.8	79.0	68090	2	AC113956
28	15.8	79.0	73397	2	AC101427
29	15.8	79.0	74475	8	TTM24
30	15.8	79.0	90632	8	AC021045
31	15.8	79.0	94319	8	ATF3L17
32	15.8	79.0	103116	2	AC128838
33	15.8	79.0	103930	9	AC092167
34	15.8	79.0	110000	2	AC119569_2
35	15.8	79.0	113721	2	AC130506
36	15.8	79.0	121542	2	AC122893
37	15.8	79.0	12697	9	AC006062
38	15.8	79.0	133279	8	OSUN00152
39	15.8	79.0	134862	2	AC116433
40	15.8	79.0	135934	2	AC018925
41	15.8	79.0	137478	2	AP004704
42	15.8	79.0	142085	2	AC036153
43	15.8	79.0	149879	2	AP005381
44	15.8	79.0	150584	2	AC106994
45	15.8	79.0	154383	2	AC097312

## ALIGNMENTS

RESULT 1	SCMRPTAA	121 bp	DNA	INV 26-APR-1993
LOCUS	SCMRPTAA			
DEFINITION	S.mansoni tandem repeat units.			
ACCESSION	M61098			
VERSION	M61098.1	GI:161064		
KEYWORDS				
SOURCE				
ORGANISM	S.mansoni (strain Egyptian) DNA, clone psm1-7.			
REFERENCE	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;			
AUTHORS	Strigeldida; Schistosomatidae; Schistosomatidae; Schistosoma.			
TITLE	1 (bases 1 to 121)			
	Hamburger,J., Turetski,T., Kapeller,I. and Deresiewicz,R.			
	Highly repeated short DNA sequences in the genome of Schistosoma			
	mansoni recognized by a species-specific probe			

```

exon          /note="myosin heavy-chain, (first expressed exon)"
              /number=2
              <2381..2564
              /note="myosin heavy-chain a, (first expressed exon)"
              /number=2
              2585..2834
              /note="MHC intron B"
              2835..2978
              /gene="myosin heavy chain"
              <2835..2978
              /gene="myosin heavy chain"
              2979..3143
              /note="MHC intron C"
              3144..3287
              /gene="myosin heavy chain a"
              3144..3287
              /partial
              /gene="myosin heavy chain a"
              3288..4137
              /note="MHC intron D"
              4138..>4518
              /note="myosin heavy-chain"
              /number=4
              4138..>4518
              /note="myosin heavy-chain a"
              /number=4
              4519..4591
              /note="MHC intron E"
              4592..>4655
              /note="myosin heavy-chain"
              /number=5
              4592..>4655
              /note="myosin heavy-chain a"
              /number=5
              4656..>4676
              /note="MHC intron F"
              1582 a 984 c 869 g 1241 t
              /origin 1 bp upstream of XbaI site.

Query Match      84.0%; Score 16.8; DB 3; Length 4676;
Best Local Similarity 90.0%; Pred. No. 2.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0,
QY      1 GATCTGAATCCGACCAACG 20
        ||||||||| 11111
Db      3106 GACTGTAATCCGACCAACG 3125

RESULT 3
DROMHC      22663 bp DNA linear INV 26-APR-1993
LOCUS      DROMHC
DEFINITION D.melanogaster myosin heavy chain gene, complete cds.
ACCESSION M61229
VERSION M61229.1 GI:157890
KEYWORDS myosin heavy chain.
SOURCE D.melanogaster (strain Canton S) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 22663)
AUTHORS George,E.L., Ober,M.B. and Emerson,C.P., Jr.
TITLE Functional domains of the Drosophila melanogaster muscle myosin
       heavy-chain gene are encoded by alternatively spliced exons
JOURNAL Mol. Cell. Biol. 9 (7), 2957-2974 (1989)
MEDLINE 89384556
PUBMED 2506434
FEATURES
source      Location/Qualifiers
            1..22663
            /organism="Drosophila melanogaster"
            /strain="Canton S"
            /db_xref="taxon:7227"
            /map="chromosome 36B(2L)"

```



```

exon      /number=12
15902..16111
/gene="MHC"
exon      /number=13
16333..17219
/gene="MHC"
exon      /number=14
17391..17469
/gene="MHC"
exon      /number=17886
17808..17886
/gene="MHC"
exon      /number=18223
18223..19398
/gene="MHC"
exon      /number=16
19466..20371
/gene="MHC"
exon      /number=17
20636..21137
/gene="MHC"
exon      /number=18
21578..22322
/gene="MHC"
exon      /number=19
21578..21771
/gene="MHC"
exon      /number=19
21753..21758
/gene="MHC"
polyA_signal
21831..21836
/gene="MHC"
polyA_signal
22304..22309
/gene="MHC"

Query Match      84.0%; Score 16.8; DB 3; Length 22663;
Best Local Similarity 90.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GATCGAATCCGACCAACCG 20
|||
Db 3108 GAACGTGATCCGACCAACCG 3127

RESULT 4
AC005119 68379 bp DNA linear INV 09-JUN-1999
DEFINITION Drosophila melanogaster, chromosome 2L, region 36B1-36B2, P1 clone
AC005119
DS06436, complete sequence.
AC005119.9 GI:5430743
HTG.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 68379)
Celinker,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chev,M., Ciesiolka,I.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Ling,H.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Richards,S., Sequiera,A., Sethi,H.,
Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M.,
Zieren,L.L. and Rubin,G.M.
Sequencing of Drosophila chromosome 2L, region 36B1-36B2
Unpublished (1998)
2 (bases 1 to 68379)
Celinker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chev,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lmotani,M.A., Mak,U., Mazda,P., Mok,M.S., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,

```

```

Sult,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
Zieren,L.L. and Kimmel,B.E.
Direct Submission
Submitted (16-JUN-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jul 9, 1999 this sequence version replaced gi:4680763.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgs@fruitfly.berkeley.edu.
P1 library location: 68-4.
Location/Qualifiers
1..68379
/organism="Drosophila melanogaster"
/strain="Y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="36B1-36B2"
/clone_lib="P1 DS06436 (D314)"
/clone_lib="P1 library, partial Sau3a in pNS582tet14d10"
/note="DS06436 (D314) extends from a minimal overlap with
its distal neighbor DS00592 (D315) at bp 1 to P1 end at bp
68378."

BASE COUNT      19786 a 15275 c 14103 g 19215 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 3; Length 68379;
Best Local Similarity 90.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GATCGAATCCGACCAACCG 20
|||
Db 44948 GAACGTGATCCGACCAACCG 44967

RESULT 5
AC019767 79309 bp DNA linear HTG 03-JAN-2000
LOCUS AC019767
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***; in ordered
pieces.
AC019767
AC019767.1 GI:6665130
VERSION AC019767.1
KEYWORDS HTG; HTGS; PHASE2.
SOURCE Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 79309)
Adams,M. and Venter,J.C.
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10210880 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..79309
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT      23312 a 16461 c 15838 g 23698 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 2; Length 79309;
Best Local Similarity 90.0%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY	1	GATCTGAATCCGACCAACCG	20
Db	74946	GAACTGAATCCGACCAACCG	74965

LOCUS	175413 bp	DNA	linear	INV 09-AUG-2001
DEFINITION	Drosophila melanogaster, chromosome 2L, region 36A-36B, BAC clone			
	BACR26003, complete sequence.			
ACCIONARY	10002007			

VERSION AC093097.1 GI:15145638  
KEYWORDS HTG.

REFERENCE	AUTHORS
1 (bases 1 to 175413)	Celniker, S.E., Adams, M.

ROGERS, Y., AN, H., BALDWIN, D., BANZON, J., BEESON, K. Y., BUSAM, D. A., CALSON, W. M., CENTER, A., CHAMPE, M., DAVEPORT, L. B., DIETZ, S. M., DODSON, K., DORSETT, V., DOUN, L. E., DOYLE, C., DRESNKE, D., FARLAN, D., FERRIERA, S., FRISKE, E., GALLE, R. F., GARY, N. S., GEORGE, R. A., GONZALEZ, M., HOUNK, J., HOSKINS, R. A., HOWLAND, T. J., IBOGWAM, C., JATAHI, M., KRUSE, D., LI, P., MATTEI, B., MOSHEFI, A., MCINTOSH, T. C., MOY, M., MURPHY, B., NELSON, C., NELSON, K. A., NUNOO, J., PACIEL, J. F., PARAGAS, V., PARK, S., PATEL, S., PEIFFER, B., PHOUANENGWONG, S., PLITMAN, G. S., PULI, Y., RICHARDS, S. J., SCHEELER, F., SWISPLETON, M., STRONG, R., SVIRSKAS, R., TECTOR, C., WILLIAMS, S. M., ZAVERI, J. S., SMITH, H. O., RUBIN, G. M. and VENTER, J. C.

REFERENCE	AUTHORS
2 (bases 1 to 175413)	
Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gockayne, T.D., Amaratilake, P.G., Branton, R.C., Rogers, Y., An, H., Baldwin, D., Baron, D., Beeson, K.Y., Busan, D.A., Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, Y., Doup, L.E., Doyle, C., Dresnak, D., Fafan, D., Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houch, S., Hoskins, R.A., Hostin, D., Howland, T.J., Ibeagwam, C., Jattali, M., Kruse, D., Li, P., Mattei, B., Mosheiff, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nimoo, J., Paclab, J., Pargass, Y., Park, S., Patel, S., Pfeiffer, B., Phoonanamong, S., Plitman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.	
Direct Submission	
Submitted (09-AUG-2001)	Berkeley Drosophila Genome Project, MS

FEATURES	COMMENT
source	64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US
	Sequence submitted by: Berkeley Drosophila Genome Project
	Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720
	This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site ( <a href="http://www.fruitfly.org/sequence/">http://www.fruitfly.org/sequence/</a> ) or send email to <a href="mailto:bdg@fruitfly.berkeley.edu">bdg@fruitfly.berkeley.edu</a> .
	Location/Qualifiers
	1..175413

```
organism="Drosophila melanogaster"  
strain="Y; cn bw sp"  
db_xref="taxon:7227"  
chromosome="2L"  
map="66A-365"  
clone="BACR6003 (DI298)"  
clone_id="PFCI-98 (Roswell Park Cancer Institute  
Library BAC library, partial EcoRI in
```

BASE COUNT	52084	a	36665	c	35604	g	51060	t	
ORIGIN	pbAc3.6j"								
Query Match:	84.0%;	Score	16.8;	DB	3;	Length	175413;		
Best Local Similarity	90.0%;	Pred. NO.	2.6e+02;						
Matches	18;	Conservative	2;	Mismatches	0;	Gaps	0;		

```

0Y      1 GATCTGATCCGACCAACCG 20
        || ||||| ||||| |||||
Db 142511 GACTGATCCGACCAACCG 142530

```

RESULT	7
Locus	AEO03652
DEFINITION	Drosophila melanogaster genomic scaffold 14200001386055 section 4
ACCESSION	AEO03652
VERSION	AEO02690
KEYWORDS	HTG.

ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 258166)

REFERENCE  
AUTHORS  
Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,  
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,  
George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,  
Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,  
Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D.,  
Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, M.,  
Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,  
Anderson, P., Antkovich, C., Baldwin, D., Baller, R.M., Baeu, A.,

Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borokov, D., Borchan, M.K., Bouck, J., Brokstein, P., Brottier, P., Burts, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L., Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, B., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferrara, C., Ferreira, S., Fleischmann, W., Foster, C., Garfield, A.E., Gary, N.S., Gelbart, W.M., Glasser, K., Gloeckl, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.F.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Idegami, C., Jalali, M., Kalusi, F., Kapen, G.H., Ke, Z., Kemison, J.A., Ketchum, K.A., Kimmel, B.E., Kodita, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasco, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,

McLeod, M. P., McNeill, A., Warkentin, G., Kilshtadt, M. V., Muehl, J. C., Morris, J., Mosnig, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nussken, D. R., Paclet, J. M., Palazzolo, M., Piltman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, R., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skipski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wasserman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, H. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M. and Venter, J. C.

TITLE The genome sequence of *Urosalpinx maculogaster*  
JOURNAL Science 287 (5461), 2185-2195 (2000)  
MEDLINE 20196006  
PMID 10731132  
2 (bases 1 to 258166)  
REFERENCE Adams,M.D., Celisliker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J  
AUTHORS Direct Submission  
TITLES Submitted (21-Mar-2000) Celera Genomics, 45 West Gude Drive,  
JOURNAL



QY 1 GATCTGAATCCGACCACCG 20  
|| ||||| |||||  
Db 153544 GAACTGAATCCGACAAACCG 153563

RESULT 8					
AC095629					
LOCUS	AC095629	99542 bp	DNA	linear	HTG 11-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-8f10, *** SEQUENCING IN PROGRESS ***				
ACCESSION	AC095629	50 unordered pieces.			
VERSION	AC095629.3	GI:21722685			
KEYWORDS	HTG; HTGS_PHASE1.				
SOURCE	Norway rat.				

REFERENCE	
AUTHORS	
1	(bases 1 to 99542)
Muzny, D.M.,	Adams, C.,

		99542 bp	DNA	linear	HTG 11-JUL-2002
AC0956629	Rattus norvegicus clone CH230-8f10,	***	SEQUENCING IN PROGRESS ***		
50 unordered pieces.					
AC0956629					
AC0956629.3	GI:21722685				
HTG; HTGS_PHASEL.					
Norway rat.					
Rattus norvegicus					
Eukaryota; Metazoa;					
Mammalia; Eutheria; Rodentia;					
Sciurognathi; Muridae; Murinae;					
Rattus					

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 99542)

REFERENCE 3 (bases 1 to 99542)

**TITLE** Direct Submission  
**JOURNAL** Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

**COMMENT**

Baylor Plaza, Houston, TX 77030, USA  
On Jul 10, 2002 this sequence replaced gi:17942183

Genome Center  
for Cell and

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

Center clone name: CH230-8F10

## ----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 43940 bases at least

Consensus quality: 47263 bases at least 0.30

Consensus quality: 50325 bases at least 030

Consensus quality: 50225 passes at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 50 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of  $N$ , but the exact sizes of the gaps are unknown

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

*	1	1745:	contig of 1745 bp in length
*	1746	1845:	gap of unknown length
*	1846	3299:	contig of 1454 bp in length
*	3300	3399:	gap of unknown length
*	3400	4667:	contig of 1268 bp in length
*	4668	4767:	gap of unknown length
*	4768	6262:	contig of 1495 bp in length
*	6263	6362:	gap of unknown length
*	6363	7679:	contig of 1317 bp in length
*	7680	7779:	gap of unknown length
*	7780	9022:	contig of 1243 bp in length
*	9023	9122:	gap of unknown length
*	9123	10340:	contig of 1218 bp in length
*	10341	10440:	gap of unknown length
*	10441	11504:	contig of 1064 bp in length
*	11505	11604:	gap of unknown length
*	11605	13075:	contig of 1471 bp in length
*	13076	13175:	gap of unknown length
*	13176	14566:	contig of 1391 bp in length
*	14567	14666:	gap of unknown length
*	14667	16154:	contig of 1488 bp in length
*	16155	16254:	gap of unknown length
*	16255	17749:	contig of 1495 bp in length
*	17750	17849:	gap of unknown length
*	17850	19505:	contig of 1656 bp in length
*	19506	19605:	gap of unknown length
*	19606	21176:	contig of 1571 bp in length
*	21177	21276:	gap of unknown length
*	21277	22846:	contig of 1570 bp in length
*	22847	22947:	gap of unknown length
*	22948	24186:	contig of 1240 bp in length
*	24187	24286:	gap of unknown length
*	24287	25953:	contig of 1669 bp in length
*	25956	26055:	gap of unknown length
*	26056	27445:	contig of 1390 bp in length
*	27446	27545:	gap of unknown length
*	27546	28570:	contig of 1025 bp in length
*	28571	28670:	gap of unknown length
*	28671	29844:	contig of 1174 bp in length
*	29844	29944:	gap of unknown length
*	29945	31571:	contig of 1627 bp in length
*	31572	31671:	gap of unknown length
*	31672	33601:	contig of 1930 bp in length
*	33602	33701:	gap of unknown length
*	33702	34808:	contig of 1107 bp in length

```

* 34809 34908: gap of unknown length
* 34909 36643: contig of 1735 bp in length
* 36644 36743: gap of unknown length
* 36744 38374: contig of 1631 bp in length
* 38375 38474: gap of unknown length
* 38475 40687: contig of 2213 bp in length
* 40688 40787: gap of unknown length
* 40788 42053: contig of 1266 bp in length
* 42054 42153: gap of unknown length
* 42154 43382: contig of 1229 bp in length
* 43383 43482: gap of unknown length
* 43483 45071: contig of 1589 bp in length
* 45072 45171: gap of unknown length
* 45172 47004: contig of 1833 bp in length
* 47005 47104: gap of unknown length
* 47105 48451: contig of 1347 bp in length
* 48452 48551: gap of unknown length
* 48552 49892: contig of 1341 bp in length
* 49893 49992: gap of unknown length
* 49993 51860: contig of 1868 bp in length
* 51861 51960: gap of unknown length
* 51961 53978: contig of 2019 bp in length
* 53980 54079: gap of unknown length
* 54080 55641: contig of 1562 bp in length
* 55642 55741: gap of unknown length
* 55742 58182: contig of 2441 bp in length
* 58183 58282: gap of unknown length
* 58283 61300: contig of 3018 bp in length
* 61301 61400: gap of unknown length
* 61401 62816: contig of 1416 bp in length
* 62817 62916: gap of unknown length
* 62917 64944: contig of 2027 bp in length
* 64944 65044: gap of unknown length
* 65044 67311: contig of 2268 bp in length
* 67312 67411: gap of unknown length
* 67412 69802: contig of 2391 bp in length
* 69803 69902: gap of unknown length
* 69903 73494: contig of 3592 bp in length
* 73495 73594: gap of unknown length
* 73595 76336: contig of 2942 bp in length
* 76337 76636: gap of unknown length
* 76637 79951: contig of 3315 bp in length
* 79952 80051: gap of unknown length
* 80052 82994: contig of 2943 bp in length
* 82995 83094: gap of unknown length
* 83095 85941: contig of 2847 bp in length
* 85942 86041: gap of unknown length
* 86042 88827: contig of 2786 bp in length
* 88828 88927: gap of unknown length
* 88928 92287: contig of 3360 bp in length
* 92288 92387: gap of unknown length
* 92388 94886: contig of 2499 bp in length
* 94887 94986: gap of unknown length
* 94987 99542: contig of 4556 bp in length.

```

## FEATURES

## Source

```

1. 99542
/db_xref="taxon:10116"
/clone="CH230-8P10"
BASE COUNT 26342 a 21077 c 19949 g 27095 t 5079 others
ORIGIN

```

## Query Match

```

Best Local Similarity 82.08; Score 16.4; DB 2; Length 99542;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 ATCTGATCCGACCAACC 19
|||||
Db 83403 ATCTGATCCGACCAACC 83420

```

```

RESULT 9
AC079890

```

```

LOCUS AC079890 132927 bp DNA linear PLN 29-AUG-2001
DEFINITION Oryza sativa chromosome 10 BAC OSJNB0089A17 genomic sequence,
ACCESSION AC079890
VERSION AC079890.13 GI:13491223
KEYWORDS HTG.
SOURCE
ORGANISM Oryza sativa.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriaraloideae; Oryzoideae; Oryza.
1 (bases 1 to 132927)
Fraser,C.M.
Oryza sativa chromosome 10 BAC OSJNB0089A17 genomic sequence
Unpublished
2 (bases 1 to 132927)
Buell,R.
Direct Submission
Submitted (16-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
3 (bases 1 to 132927)
Buell,R.
Direct Submission
Submitted (31-MAR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
4 (bases 1 to 132927)
Buell,R.
Direct Submission
Submitted (28-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rpuell@tigr.org
5 (bases 1 to 132927)
Buell,R.
Submitted (10-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rpuell@tigr.org
6 (bases 1 to 132927)
Buell,R.
Submitted (29-AUG-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rpuell@tigr.org
On Mar 31, 2001 this sequence version replaced gi:13310897.
Address all correspondence to: rpuell@tigr.org
BAC clone OSJNB0089A17 is from Oryza sativa chromosome 10
The orientation of the sequence is from Sp6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including GENSCAN and GENSCAN+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GENMARKHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/genemark/), Egenesh
(http://www.softberry.com/), and GeneSplicer (Michael Perle and
Steven Salzberg, contact mperle@tigr.org), searches of the
complete sequence against a peptide database and the Plant EST
database at TIGR (http://www.tigr.org/db/tgi.shtml). Annotated
genes are named to indicate the level of evidence for their
annotation. Genes with similarity to other proteins are named after
the database hits. Genes without significant peptide similarity but
with EST similarity are named as unknown proteins. Genes without
protein or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are predicted as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Rddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES
Source
1. 132927
/organism="Oryza sativa"
/cultivar="Nipponbare"
/sub_species="japonica"

```

repeat\_region /db\_xref="taxon:4530"  
/map="near C797"  
/clone="OSJNB0089A17"  
complement(3708..3735)  
/rpt\_family="AT\_rich"  
repeat\_region complement(8219..8252)  
/rpt\_family="(GAAA)n"  
complement(10188..10210)  
repeat\_region /rpt\_family="AT\_rich"  
11133..11176  
/rpt\_family="(GAA)n"  
11215..11237  
repeat\_region /rpt\_family="(GAAA)n"  
complement(11764..11796)  
repeat\_region /rpt\_family="AT\_rich"  
11867..11918  
gene /rpt\_family="(GAAA)n"  
complement(11962..12830)  
/gene="OSJNB0089A17.22"  
mrna /note="Similar to transcriptional activator CBFI  
GB:A4C4962 GI:1899058 (Arabidopsis thaliana)"  
complement(join<11962..12330,12423..>12830)  
/gene="OSJNB0089A17.22"  
cds complement(join11962..12330,12423..12830)  
/gene="OSJNB0089A17.22"  
/codon\_start=1  
/product="putative transcriptional factor"  
protein /protein\_id="AAK31271.1"  
/db\_xref="GI:13569987"  
/translation="YAMAKKELOETSSSSSSNAATSSCSAAVTDAWSSPARNPAAVAG  
KKKKEVGEADAEAGAGEEEEAARSAARSSAKRKRSQKHPPYRGVRR  
AGKKEVSEIREPEKSRINIGTPTDMMARAHDDVQAAALAAATSPSSSSSL  
SNDDVACVVAHADDEPRAAAAKKDDDDSTTAPATAAAADADQQLPDLPLFLD  
IODEPGRPAMAPLADVDYNAETLRERPLINDLGLTDA"  
repeat\_region complement(12126..12162)  
/rpt\_family="(CGG)n"  
repeat\_region complement(12565..12703)  
/rpt\_family="(GGG)n"  
12750..12802  
repeat\_region /rpt\_family="(GGA)n"  
complement(13363..13409)  
repeat\_region /rpt\_family="AT\_rich"  
complement(13435..13464)  
repeat\_region /rpt\_family="AT\_rich"  
complement(15206..15247)  
repeat\_region /rpt\_family="AT\_rich"  
complement(15414..15465)  
repeat\_region /rpt\_family="AT\_rich"  
complement(15587..16007)  
/gene="OSJNB0089A17.21"  
/note="Predicted by glimmer"  
mrna complement(>15582..>16007)  
/gene="OSJNB0089A17.21"  
cds complement(15582..16007)  
/gene="OSJNB0089A17.21"  
/codon\_start=1  
/product="hypothetical protein"  
protein /protein\_id="AAK31268.1"  
/db\_xref="GI:13569984"  
/translation="MDYCEPPSKPGVALAPMLSPGHVTQITAAITRPSVORTSR  
AMAAGACVADSESHGVGRCERVRYHETLRARODDARRAALRSIDROYACID  
RSPPKRKAYDGLSVYGLHSYDQPKTHAFYLRKFWAGA"  
repeat\_region complement(17572..17613)  
repeat\_region /rpt\_family="(CGG)n"  
complement(17622..17642)  
repeat\_region /rpt\_family="GC\_rich"  
17938..21425  
gene /gene="OSJNB0089A17.1"  
/note="Similar to 8 amino 7 oxononate synthase  
GB:J00512 GI:98180 (Bacillus spheericus); EST AU094796  
from this gene"  
mrna join<17938..18133,18368..18480,18672..18779,18871..19109,

```

1952. 19439,19521. 19654,20178. 20266,20622. 20655,
20734. 20802,21120. 21425)
/gene="OSJNB0089A17.1"
join(17998, 18133,18366, .18480,18672, .18779,18871, .19109,
19352. 19439,19521. 19654,20178. 20266,20622. 20655,
20734. 20802,21120. 21243)
/gene="OSJNB0089A17.1"
/codon_start=1
/product="putative 8-amino-7-oxononanoate synthase"
/protein_id="AAK31282.1"
/db_xref="GI:13569988"
/translation="MAQMDALVDNALALRLASRLIRPRRLAAPPAPPPPCPP
WDAAVEITLDRLTHQLALRGEGSGGGEKDEKLLIFSGNDYMLSSHPIRLHAY
KAEEYGGKPPRGSAALICGTYTHRWAEELSLAKKEDCLCPFGSAMAAVMLAN
ISSILAGKGRPAEDERIALISDALINHASITDGLRLVERQOEYVSFYKICDMSHLL
LLTSMGDGAPLPELTVLRRKYGFLLVIDADGTLVCGENGGAPLPECEMEDISY
GLTSKAGCGGGEGTICSTWKKRLIOSRGSEFTSLALPVYVAAYVSGEKAALRAG
DMRSGFHTPIRPPTVPVPSNCRITLTSASSDDILRLVDALTFWMLPHKDDIITA
VASKL"
complement(18024, .18046)
/rpt_family="GC-rich"
24913. 24939
/rpt_family="(CAGTT)n"
complement(24934, .26046)
/gene="OSJNB0089A17.20"
/note="predicted by figenes"
complement(join(24934, .25129,25223, .>26046))
/gene="OSJNB0089A17.20"
complement(join(24934, .25129,25223, .26046))
/gene="OSJNB0089A17.20"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK31265.1"
/db_xref="GI:13569981"
/translation="MASGAVALLMFTHTERELTFRVLVGHGGQPPPMRWYIALMLT
ESVGHDFPRVAVLPAPVLRVVDALCLARLRPRRGVAGGAERRIALAAAGDAD
PALRLPCNTNALAEYVEGSLATPDARDEVMYGVSDYRVNVCRIYEDDGVAAVAAD
DDDAALAAATLPRVDLALDGTTPRRPRPRRYHOYHNHYNAMAPLPPRYAELNPA
ASFWPEYQOEQPPRPPOOHGVIPLREYSLSTFTFSRQYPIRODILINFNSL
YGCVCVSVVVEKAAAGQIPYGRVYLRCPSMLPVLDGQTAHYMKSHMARLYVP
SRPN"
25332. 25357
/rpt_family="(CGG)n"
25414. 25502
/rpt_family="(CGG)n"
25541. 25594
/rpt_family="(CGG)n"
complement(25734, .25822)
/rpt_family="GC-rich"
29995. 32244
/gene="OSJNB0089A17.10"
/note="similar to quinone oxidoreductase GB:AAA23691
GI:145766 [Escherichia coli]"
join(29995, .30039,30157, .30295,31110, .31336,31482, .31739

```



```
misc_feature 15165..22781
                /note="assembly_fragment"
misc_feature 22882..39817
                /note="assembly_fragment"
misc_feature 39918..58842
                /note="assembly_fragment"
misc_feature 58943..102439
                /note="assembly_fragment"
                clone_end:17
                vector_side:left
                102540..155813
misc_feature /note="assembly_fragment"
BASE COUNT 51921 a 30751 c 29180 g 43159 t 802 others
ORIGIN
Query Match 82.0%; Score 16.4; DB 2; Length 155813;
Best Local Similarity 94.4%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCGAATCCGACCAAC 18
Db 1559 GATCGAATCCTACCAAC 1542

RESULT 12
LOCUS AC091914 165237 bp DNA linear PRI 27-FEB-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-221N1, complete sequence.
ACCESSION AC091914
VERSION AC091914.3 GI:18959840
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 165237)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Unpublished
AUTHORS 2 (bases 1 to 165237)
JOURNAL DOE Joint Genome Institute.
REFERENCE Direct Submission
AUTHORS Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 165237)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (27-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 27, 2002 this sequence version replaced gi:14579738.
Draft Sequence Produced by DOE Joint Genome Institute
WWW.JGI.DOE.GOV
Finishing Completed at Stanford Human Genome Center
WWW.SHGC.STANFORD.EDU
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated total Number of Errors is 0.5.
FEATURES
source 1..165237
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RP11-221N1"
BASE COUNT 46852 a 30135 c 32136 g 56114 t
ORIGIN
Query Match 82.0%; Score 16.4; DB 9; Length 165237;
Best Local Similarity 94.4%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCGAATCCGACCAAC 18
Db 37737 GATCGAATCCTACCAAC 37754
```

```
RESULT 13
LOCUS AC114308 177246 bp DNA linear HTG 07-MAR-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-197F21, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
ACCESSION AC114308
VERSION AC114308.1 GI:19224957
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 177246)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
REFERENCE Sequencing of Human Chromosome 5
AUTHORS 2 (bases 1 to 177246)
JOURNAL DOE Joint Genome Institute.
COMMENT Direct Submission
Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 482154
Center clone name: RPCR-11_197F21
-----
Summary Statistics
Consensus quality: 171493 bases at least Q40
Consensus quality: 173924 bases at least Q30
Consensus quality: 174945 bases at least Q20
Estimated insert size: 175000; agarose-ff estimation
Estimated insert size: 176846; sum-of-contigs estimation
Quality coverage: 9.88 in Q20 bases; agarose-ff estimation
Quality coverage: 9.77 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1270: contig of 1270 bp in length
* 1271 1370: gap of unknown length
* 1272 12248: contig of 10878 bp in length
* 1249 12348: gap of unknown length
* 12349 33985: contig of 21637 bp in length
* 33986 34085: gap of unknown length
* 34086 79299: contig of 45214 bp in length
* 79300 79399: gap of unknown length
* 79400 177246: contig of 97847 bp in length.
FEATURES
source 1..177246
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RP11-197F21"
        /clone_lib="RP11 human BAC library 11"
BASE COUNT 54617 a 32530 c 33575 g 56124 t 400 others
ORIGIN
Query Match 82.0%; Score 16.4; DB 2; Length 177246;
Best Local Similarity 94.4%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCGAATCCGACCAAC 18
Db 84991 GATCGAATCCTACCAAC 85008
```

RESULT 14  
AC018248/c  
LOCUS AC018248  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
AC018248 9359 bp DNA linear HTG 09-DEC-1999  
AC018248.1 GI:6552943  
VERSION HTG: HTGS\_PHASE2.  
KEYWORDS Drosophila melanogaster.  
SOURCE Drosophila melanogaster.  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 9359)  
AUTHORS Adams, M. and Venter, J.C.  
TITLE Direct Submission  
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
COMMENT This sequence was identified as CDM:10214038 by the submitter. For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1. 9359  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 2737 a 2174 c 1909 g 2539 t  
ORIGIN

Query Match 80.0%; Score 16; DB 2; Length 9359;  
Best Local Similarity 100.0%; Pred. No. 7.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATCTGAATCCGACCAA 17  
|||||  
Db 7973 ATCTGAATCCGACCAA 7958

RESULT 15  
AC012160/c  
LOCUS AC012160  
DEFINITION Drosophila melanogaster, chromosome X, region 15E-15E, BAC clone BACR06602, complete sequence.  
AC012160  
VERSION AC012160.6 GI:17646875  
KEYWORDS HTG.  
SOURCE Drosophila melanogaster.  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 172069)  
AUTHORS Celisner, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busan, D.A., Carlson, J.W., Center, A., Chame, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, Y., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Fitze, E., Galle, R.F., Gary, N.S., George, R.A., Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J., Idegawa, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Sylvestre, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.  
TITLE Sequencing of Drosophila chromosome X, region 15E-15E  
JOURNAL Unpublished (1998)  
REFERENCE 2 (bases 1 to 172069)  
AUTHORS Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,

Butenhoff, C., Chame, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Iomutan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequiera, A., Settl, H., Sili, E., Sylvestre, R.R., Wan, K.H., Weindburg, T., Zhang, R., Zieran, L.T. and Rubin, G.M.  
TITLE Direct Submission  
JOURNAL Submitted (21-OCT-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
COMMENT On Dec 13, 2001 this sequence version replaced gi:6466928.  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [hdg@fruitfly.berkeley.edu](mailto:hdg@fruitfly.berkeley.edu).  
Location/Qualifiers  
1. 172069  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="X"  
/map="15E-15F"  
/clone="BACR06602 (D1110)"  
/clone\_id="RPcl-98 (Roswell) Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6")  
BASE COUNT 48971 a 35590 c 36029 g 51479 t  
ORIGIN

Query Match 80.0%; Score 16; DB 3; Length 172069;  
Best Local Similarity 100.0%; Pred. No. 7.2e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATCTGAATCCGACCAA 17  
|||||  
Db 162241 ATCTGAATCCGACCAA 162226

Search completed: May 12, 2003, 01:51:34  
Job time : 391 secs

biological sample especially in cases of low infection

PT Schistosoma by polymerase chain reaction -

```
XX Claim 4, Page 13, 34pp; English.
PS
XX The present sequence is that of a claimed oligonucleotide primer,
CC derived from the 5' region of a highly repeated short DNA sequence
CC (see AAI70400) of the Schistosoma mansoni genome. This 5' primer,
CC and the 3' primer given in AAI70402, are used in the method of the
CC invention for diagnosing Schistosoma infection. The method involves
CC collection of the sample to be examined, extraction of Schistosoma
CC sp. DNA, amplification by PCR, separation of PCR products by
CC electrophoresis, and detection e.g. by colouring with silver salts.
CC A kit for diagnosing infection, which includes the primers, is
CC claimed. The method is useful for detecting Schistosoma sp.
CC parasites by detecting the parasite's DNA in a biological sample.
CC It is especially useful in cases of low infection intensity for
CC which parasitological stool tests demonstrate little sensitivity.
SQ Sequence 20 BP; 6 A; 7 C; 4 G; 3 T; 0 other;

Query Match          100.0%; Score 20; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATCTGATCCGACCAACG 20
Db 1 GATCTGATCCGACCAACG 20

RESULT 2
AAI70400
ID AAI70400 standard; DNA; 120 BP.
AC AAI70400;
XX
XX 07-JAN-2002 (first entry)
DE Schistosoma mansoni repeater unit DNA.
XX Schistosomiasis; infection; diagnosis; ds.
KW Schistosoma mansoni.
OS
XX
XX key Location/Qualifiers
XX primer_bind complement (1..19)
XX /tag= a
XX /note= "Primer of AAI70401"
XX primer_bind 91..110
XX /tag= b
XX /note= "Primer of AAI70402"
XX
XX WO200175148-A1.
XX
XX 11-OCT-2001.
XX
XX 04-APR-2001; 2001WO-BR00035.
XX
XX 04-APR-2000; 2000BR-0001536.
XX
XX (FIOC-) FIOCRUZ FUNDACAO CRUZ OSWALDO.
XX
XX Teles Rabello AL; Dias Neto E; Pontes LA;
XX
XX WPI; 2001-648561/74.
XX
XX Diagnosing infection by parasites of the Schistosoma sp. in a
XX biological sample especially useful in cases of low infection
XX intensity, comprises detecting a specific region of the DNA of
XX Schistosoma by polymerase chain reaction -
XX
XX Claim 1; Page 22; 34pp; English.
XX
XX The present sequence is that of a highly repeated short DNA
XX sequence in the genome of Schistosoma mansoni. PCR primers
```

```
CC (see AAI70401 and AAI70402) that flank this sequence are used in
CC the method of the invention for diagnosing Schistosoma infection.
CC The method involves collection of the sample to be examined,
CC extraction of Schistosoma sp. DNA, amplification by PCR, separation
CC of PCR products by electrophoresis, and detection e.g. by colouring
CC with silver salts. A kit for diagnosing infection, which includes
CC the primers, is claimed. The method is useful for detecting in a
CC Schistosoma sp. parasites by detecting the DNA of the parasite in a
CC biological sample. It is especially useful in cases of low
CC infection intensity for which parasitological stool tests
CC demonstrate little sensitivity.
SQ Sequence 120 BP; 36 A; 20 C; 26 G; 38 T; 0 other;

Query Match          100.0%; Score 20; DB 22; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATCTGATCCGACCAACG 20
Db 1 GATCTGATCCGACCAACG 20

RESULT 3
AAA46170
ID AAA46170 standard; DNA; 6494 BP.
AC AAA46170;
XX
XX 27-SEP-2000 (first entry)
DE
DE GFP-Sm1-7 fusion protein construct.
DE
XX
XX GFP-Sm1-7 fusion construct; circular; green fluorescent protein;
XX glutathione S-transferase; eukaryotic diploid multicellular parasite;
XX universal graft; transgenic eukaryotic parasite; acquired deficiency;
XX genetic deficiency; hormone deficiency; metabolic deficiency;
XX haematological deficiency; immunological deficiency; immunotherapy;
XX anti-microbial therapy; anti-cancer therapy; drug addiction;
XX poisoning condition; geriatric condition; ds.
XX
XX Synthetic.
XX
XX WO2000032804-A1.
XX
XX 08-JUN-2000.
XX
XX 01-DEC-1999; 99WO-1100651.
XX
XX 01-DEC-1998; 98US-0201850.
XX
XX (YISS ) YISSUM RES & DEV CO.
XX
XX Hamburger J; Iabon A;
XX
XX WPI; 2000-412348/35.
XX
XX Eukaryotic diploid multicellular parasite useful as universal grafts
XX for in vivo delivery of beneficial gene products in humans and animals
XX involves transformation with a transgene -
XX
XX Example 2; Fig 6; 90pp; English.
XX
XX This sequence represents a GFP-Sm1-7 fusion construct contained within a
XX recombinant vector. This sequence contains the green fluorescent protein
XX (GFP) coding sequence from Aequorea victoria, and the
XX Sm1-7 coding sequence from Schistosoma, along with
XX promoter sequences. The invention relates to a eukaryotic diploid
XX multicellular parasite transformed with a transgene. Transgenic
XX eukaryotic parasites are useful as universal grafts for in vivo delivery
XX of beneficial gene product in humans and animals. The parasites can
XX particularly be used for restoration of deficiencies whether acquired or
XX genetic, such as hormone deficiencies, metabolic deficiencies,
```

CC haematological deficiencies, immunological deficiencies, immunotherapy,  
 CC anti-microbial therapy, anti-cancer therapy. They can also be used for  
 CC treatment of drug addiction, of poisoning conditions, and for  
 CC amelioration of geriatric conditions. Treatment of humans with in vivo  
 CC transgenes are universally compatible, readily available and inexpensive.  
 CC Genotypic alterations of the patients is avoided, reducing risks of  
 CC mutagenesis and malignant transformation.  
 XX

Sequence 6494 BP; 1754 A; 1373 C; 1481 G; 1886 T; 0 other;

Query Match 100.0%; Score 20; DB 21; Length 6494;  
 Best Local Similarity 100.0%; Pred. No. 1.6;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTGAATCCGACCAACCG 20  
 |||||

Db 3055 GATCTGAATCCGACCAACCG 3074

RESULT 4  
 AAT83836

ID AAT83836 standard; DNA; 467 BP.

XX AAT83836;

DT 20-AUG-1998 (first entry)

DE DNA encoding a Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;

KW Staphylococcal gene; regulatory element; bacterial gene expression;

KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;

XX toxic shock syndrome; ss.

OS Staphylococcus aureus.

XX Staphylococcus aureus.

FT Key Location/Qualifiers

FT CDS 144..419

PN W09730070-A1.

XX 21-AUG-1997.

PF 19-FEB-1997; 97WO-US03318.

XX 20-FEB-1996; 96US-0011888.

PR (SMK ) SMITHKLINE BEECHAM CORP.

PI Black MT, Burnham MK, Hodgson JE, Knowles DVC, Nicholas RO;

PI Pratt JM, Reichard RM, Rosenberg M, Ward JM;

DR WPI; 1997-424969/39.

XX P-PSDB; AAW27874.

PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used

PT to isolate antimicrobial compounds, and in vaccines against S.

XX aureus infection

PS Claim 9; Page 702; 989pp; English.

XX The present sequence encodes a Staphylococcus aureus protein of  
 CC unknown function. The present sequence was isolated from a  
 CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA  
 CC sequence can be used in the construction of ribozymes and antisense  
 CC sequences to control the expression of Staphylococcal genes. The DNA  
 CC sequence is also useful as a source of regulatory elements for the  
 CC control of bacterial gene expression. The encoded protein may be used  
 CC to produce vaccines to enable a host to produce specific antibodies  
 CC with antibacterial action. These vaccines and antibodies would protect  
 CC a host against invasion by S. aureus, and conditions relating to  
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled

CC skin syndrome, and toxic shock syndrome.

XX Sequence 467 BP; 192 A; 83 C; 86 G; 99 T; 7 other;

Query Match 84.0%; Score 16.8; DB 18; Length 467;

Best Local Similarity 90.0%; Pred. No. 36;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGAATCCGACCAACCG 20  
 |||||

Db 34 GATCTGAATCCGACCAACCG 53

RESULT 5  
 ABU28616

ID ABU28616 standard; DNA; 22188 BP.

XX ABU28616;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37321.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

PN W0200171042-A2.

XX 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

XX Claim 1; SEQ ID NO 37321; 21pp + sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA

CC sequences (ABU1840-ABU16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 22188 BP; 6664 A; 5376 C; 4559 G; 5589 T; 0 other;

Query Match 84.0%; Score 16.8; DB 23; Length 22188;

Best Local Similarity 90.0%; Pred. No. 71;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGAATCCGACCAACCG 20  
 |||||

Db 1843 GATCTGAATCCGACCAACCG 1862

RESULT 6

XX Zea mays.  
PN WO8624796-A1.  
XX  
PD 11-JUN-1998.  
XX  
PF 01-DEC-1997; 97WO-US21782.  
PR 07-MAR-1997; 97US-0813507.  
PR 02-DEC-1996; 96US-0032069.  
XX  
PA (AFFY-) AFFYMETRIX INC.  
PI Landry BS, Lemieux B, Murgineux A, Sapolsky RJ;  
PI WPI; 1998-333252/29.  
XX  
DR Brassica species allele-specific oligonucleotide probes and primers  
XX - useful for plant breeding  
PT  
PT  
PS Example 1; Page Page 53; 65pp; English.  
XX  
XX AA51401-V51704 are forward PCR primers used to amplify fragments of the  
CC zea mays genome in order to detect polymorphic markers. Such markers can  
CC be used in the construction of allele-specific primers and probes for  
CC identification or hybridisation, e.g. to determine common or disparate  
CC ancestry between 2 or more plants, to monitor the genetic contribution  
CC of an ancestral plant, to trace the progeny of proprietary plants, in  
CC certification of a hybrid plant or to identify the progeny of a  
CC back-crossed plant with an ancestral plant.  
XX  
SQ Sequence 20 BP; 5 A; 8 C; 3 G; 4 T; 0 other;  
XX  
Query Match 79.0%; Score 15.8; DB 19; Length 20;  
Best Local Similarity 89.5%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 GATCTGATTCGACCACAC 19  
|||11111111111111111111  
DB 1 GATCTGATTCGACCACAC 19  
XX  
RESULT 8  
AAK11862  
ID AAK11862 standard; DNA; 530 BP.  
AC AAK11862;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SPQ ID NO: 11853.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
PD 09-AUG-2001.  
PD 30-JAN-2001; 2001WO-US00667.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.

```

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 11853; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 530 BP; 151 A; 132 C; 59 G; 188 T; 0 other;
XX
XX Query Match 79.0%; Score 15.8; DB 22; Length 530;
XX Best Local Similarity 89.5%; Pred. No. 1.8e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 2 ATCTGATTCGACCAACCG 20
XX |||||||||||||
XX 2 ATCTGATTCGACCAACTG 20
XX
XX
XX RESULT 9
XX ID AAS5514/C
XX AAS5514 standard; DNA; 1005 BP.
XX
XX AAS5514;
XX
XX 13-FEB-2002 (first entry)
XX
XX Streptococcus pneumoniae DNA for cellular proliferation protein #85.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
XX Streptococcus pneumoniae.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT- ) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
XX P-PSDB; AAU37655.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 9151; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the

```

```

CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
XX ftp.wipo.int/pub/published_pcr_sequences.
XX
XX Sequence 1005 BP; 277 A; 175 C; 255 G; 298 T; 0 other;
XX
XX Query Match 79.0%; Score 15.8; DB 23; Length 1005;
XX Best Local Similarity 89.5%; Pred. No. 1.9e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 2 ATCTGATTCGACCAACCG 20
XX |||||||||||||
XX 596 ATCTGATTCGACCAACTG 578
XX
XX
XX RESULT 10
XX ID ABQ53190/C
XX ABQ53190 standard; DNA; 1173 BP.
XX
XX ABQ53190;
XX
XX 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 38781.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP10074.
XX 01-SEP-2000; 2000DE-1043826.
XX 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI: 2002-371823/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA -
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one

```

CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 1173 BP; 164 A; 142 C; 451 G; 416 T; 0 other;

Query Match 79.0%; Score 15.8; DB 24; Length 1173;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTGAATCGACCAACCG 20

DB 920 ACCTAAATCGACCAACCG 902

RESULT 11

ABQ53191

XX ABQ53191 standard; DNA; 1173 BP.

XX ABQ53191;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 39782.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX drug; side effect; cancer; central nervous system; cardiovascular;

XX gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Claim 12; 56bp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two

CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 1173 BP; 416 A; 451 C; 142 G; 164 T; 0 other;

Query Match 79.0%; Score 15.8; DB 24; Length 1173;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ATCTGAATCGACCAACCG 20

DB 254 ACCTAAATCGACCAACCG 272

RESULT 12

ABQ52080/C

XX ABQ52080 standard; DNA; 1210 BP.

XX ABQ52080;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 38671.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX drug; side effect; cancer; central nervous system; cardiovascular;

XX gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Claim 12; 56bp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

DD 3 CIGAA1CCGALCFAAGCG 41

```

RESULT 15
ABL21263
ID ABL21263 standard; DNA; 723 BP.
XX
AC ABL21263;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide seq ID NO 15262.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 15262; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABR57737-ABR72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 723 BP; 163 A; 195 C; 193 G; 172 T; 0 other;
XX
Query Match 77.0%; Score 15.4; DB 23; Length 723;
Best Local Similarity 94.1%; Pred. No. 2.9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 TGTGATCGACCAACC 19
||||| |||||||
DB 303 TGTGATCGACCAACC 319

```

Search completed: May 12, 2003, 01:14:56  
 Job time : 40.2143 secs

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:08:23 ; Search time 8.78571 Seconds  
(without alignments)  
698.126 Million cell updates/sec

Title: US-09-980-277-2  
Perfect score: 20  
Sequence: 1 gatctgaatccgaccacacg 20

Scoring table: IDENTIFY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, NA: \*  
1: /cgn2\_6/prodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCITUS.COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfillseq1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	15.8	79.0	3671 4	US-08-932-787B-5
2	15.8	79.0	3671 4	US-08-932-787B-5
3	15.8	79.0	3671 4	US-08-932-787B-5
4	15.2	76.0	1531 2	US-08-948-176-24
5	15.2	76.0	1531 2	US-08-948-176-24
6	14.8	74.0	2794 1	US-08-435-925C-1
7	14.8	74.0	4843 3	US-08-986-485-1
8	14.2	71.0	587 4	US-09-385-982-497
9	14.2	71.0	1080 4	US-09-038-219B-1
10	14.2	71.0	2916 4	US-08-976-259-22
11	14.2	71.0	3671 1	US-08-176-620A-5
12	14.2	71.0	3671 1	US-08-463-862-5
13	14.2	71.0	3671 2	US-08-461-983-5
14	14.2	71.0	3671 2	US-08-458-887-5
15	14.2	71.0	3677 2	US-08-390-888A-2
16	14.2	71.0	6496 4	US-09-221-017B-543
17	14.2	71.0	10846 4	US-09-098-219B-5
18	14.2	71.0	10900 4	US-09-098-219B-5
19	13.8	69.0	555 4	US-08-479-089A-4
20	13.8	69.0	691 4	US-08-961-527-381
21	13.8	69.0	1003 4	US-08-961-527-381
22	13.8	69.0	1168 4	US-08-961-527-378
23	13.8	69.0	1197 4	US-08-961-527-362
24	13.8	69.0	1949 4	US-09-134-001C-2332
25	13.8	69.0	2335 4	US-08-961-527-289
26	13.8	69.0	2393 4	US-08-961-527-275
27	13.8	69.0	2526 1	US-07-912-952-1

28	13.8	69.0	3763 4	US-08-961-527-186	Sequence 186, App
29	13.8	69.0	4137 4	US-09-221-017B-329	Sequence 329, App
30	13.8	69.0	4266 4	US-09-651-011A-3	Sequence 3, App1
31	13.8	69.0	6492 4	US-08-961-527-188	Sequence 188, App
32	13.8	69.0	6516 4	US-08-961-527-105	Sequence 105, App
33	13.6	68.0	30 4	US-09-545-244A-3	Sequence 3, App1
34	13.6	68.0	38 3	US-09-109-063-17	Sequence 16, App1
35	13.6	68.0	41 3	US-09-109-063-16	Sequence 17, App1
36	13.6	68.0	218 3	US-09-035-648-14	Sequence 14, App1
37	13.6	68.0	218 4	US-09-001-951-14	Sequence 14, App1
38	13.6	68.0	218 4	US-08-818-828-14	Sequence 14, App1
39	13.6	68.0	372 3	US-08-908-643C-4	Sequence 4, App1
40	13.6	68.0	827 5	PCT-US95-06406A-20	Sequence 20, App1
41	13.6	68.0	993 3	US-09-109-063-2	Sequence 17, App1
42	13.6	68.0	1176 3	US-08-689-421-18	Sequence 18, App1
43	13.6	68.0	1176 4	US-09-389-528-18	Sequence 18, App1
44	13.6	68.0	1176 4	US-09-181-827A-18	Sequence 18, App1
45	13.6	68.0	1347 3	US/08/622	INFORMATION FOR

#### ALIGNMENTS

```

RESULT 1
US-08-932-787B-5
; Sequence 5, Application US/08932787B
; Patent No. 6297963
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: ANTIBODIES DIRECTED TOWARD EXTRACELLULAR SIGNAL-RELATED
; FILE REFERENCE: REG 430-A-1
; CURRENT APPLICATION NUMBER: US/08/932,787B
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 08/469,547
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
; PRIOR APPLICATION NUMBER: 07/701,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3671
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (303)...(2018)
; OTHER INFORMATION: ERK3 CDNA
US-08-932-787B-5

Query Match          79.0%; Score 15.8; DB 4; Length 3671;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 GATCTGAATCCGACCAAC 19
||||||| ||| |||||
Db      843 GATCTGAATCCGACCAAC 861

RESULT 2
US-08-932-012C-5
; Sequence 5, Application US/08932012C
; Patent No. 6297035
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; FILE REFERENCE: REG 430-Y-1
; CURRENT APPLICATION NUMBER: US/08/932,012C
; PRIOR FILING DATE: 1997-09-17

```

PRIOR APPLICATION NUMBER: 08/462,874  
PRIOR FILING DATE: 1995-06-05  
PRIOR APPLICATION NUMBER: 08/178,488  
PRIOR FILING DATE: 1994-01-07  
PRIOR APPLICATION NUMBER: 07/701,544  
PRIOR FILING DATE: 1991-05-16  
PRIOR APPLICATION NUMBER: 07/532,004  
PRIOR FILING DATE: 1990-06-01  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5  
LENGTH: 3671  
TYPE: DNA  
ORGANISM: RAT  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (303)..(2018)  
OTHER INFORMATION: ERK3 cDNA  
US-08-932-012C-5

Query Match 79.0%; Score 15.8; DB 4; Length 3671;  
Best Local Similarity 89.5%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGATCCGACCAACC 19  
DB 843 GATCTGATCCGACCAACC 861

RESULT 3  
US-08-888-818C-5  
Sequence 5, Application US/08888818C  
Patent No. 6303358  
GENERAL INFORMATION:  
APPLICANT: Boulton et al.  
TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES  
FILE REFERENCE: REG 430-V-1  
CURRENT APPLICATION NUMBER: US/08/888,818C  
CURRENT FILING DATE: 1997-07-07  
PRIOR APPLICATION NUMBER: 08/478,985  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 08/178,488  
PRIOR FILING DATE: 1994-01-07  
PRIOR APPLICATION NUMBER: 07/701,544  
PRIOR FILING DATE: 1991-05-16  
PRIOR APPLICATION NUMBER: 07/532,004  
PRIOR FILING DATE: 1990-06-01  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5  
LENGTH: 3671  
TYPE: DNA  
ORGANISM: RAT  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (303)..(2018)  
OTHER INFORMATION: ERK3 cDNA  
US-08-888-818C-5

Query Match 79.0%; Score 15.8; DB 4; Length 3671;  
Best Local Similarity 89.5%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGATCCGACCAACC 19  
DB 843 GATCTGATCCGACCAACC 861

RESULT 4  
US-08-948-176-24  
Sequence 24, Application US/08948176  
Patent No. 5945385  
GENERAL INFORMATION:

APPLICANT: HTZ, WILLIAM D.  
APPLICANT: YADAV, NARENDRA S.  
TITLE OF INVENTION: ACYL-ACP THIOESTERASES GENES  
TITLE OF INVENTION: AND THEIR USE IN ALTERING PLANT  
TITLE OF INVENTION: OIL COMPOSITION  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95  
SOFTWARE: MICROSOFT WORD VERSION 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,176  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/631,264  
FILING DATE: DECEMBER 20, 1990  
ATTORNEY/AGENT INFORMATION:  
NAME: CHRISTENBURY, LYNN M.  
REGISTRATION NUMBER: 30,971  
REFERENCE/DOCKET NUMBER: CR-8926-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-5481  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1531 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 80..1192  
US-08-948-176-24

Query Match 76.0%; Score 15.2; DB 2; Length 1531;  
Best Local Similarity 85.0%; Pred. No. 35;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCTGATCCGACCAACC 20  
DB 1062 GATCTGATCCGACCAACC 1081

RESULT 5  
US-09-221-017B-373/C  
Sequence 373, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Montoy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 7061441
INFORMATION FOR SEQ ID NO: 373:
SEQUENCE CHARACTERISTICS:
LENGTH: 7096 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORCYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...7096
US-09-221-017B-373

Query Match 75.0%; Score 15; DB 4; Length 7096;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGAATCGACGACACC 19
DB 4931 TGAATCGACGACACC 4917

RESULT 6
US-08-435-925C-1
Sequence 1, Application US/08435925C
Patent No. 5646025
GENERAL INFORMATION:
APPLICANT: Moyer, Donna
TITLE OF INVENTION: SCYTALIDUM CATALASE GENE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 5646025 of No. 5646025th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
MEDIUM TYPE: Floppy disk
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,925C
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
```

```

ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4429.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2794 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Intron
LOCATION: 283..413
FEATURE:
NAME/KEY: Intron
LOCATION: 618..696
FEATURE:
NAME/KEY: Intron
LOCATION: 718..793
FEATURE:
NAME/KEY: Intron
LOCATION: 956..1108
FEATURE:
NAME/KEY: Intron
LOCATION: 1166..1218
FEATURE:
NAME/KEY: Intron
LOCATION: 1789..1842
FEATURE:
NAME/KEY: Intron
LOCATION: 2671..2764
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..282, 414..617, 697..717, 794..955, 1109
..1165, 1219..1788, 1843..2670, 2765..2791)
US-08-435-925C-1

Query Match 74.0%; Score 14.8; DB 1; Length 2794;
Best Local Similarity 88.9%; Pred. No. 61;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCGAATCGACCAAC 18
DB 1537 GATCGAATCGACCAAC 1554

RESULT 7
US-08-986-485-1
Sequence 1, Application US/08986485
Patent No. 6046030
GENERAL INFORMATION:
APPLICANT: WU, SHUTIAN
APPLICANT: SWEET, RAYMOND
APPLICANT: TRUENH, ALESEGED
TITLE OF INVENTION: A HUMAN LIG-1 HOMOLOG (HLIG-1)
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: RAYNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986,485
```

```

; FILING DATE: 08-DEC-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/059,448
; FILING DATE: 22-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GR-70264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4843 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-986-485-1

Query Match
74.0%; Score 14.8; DB 3; Length 4843;
Best Local Similarity 88.9%; Pred. No. 66;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCGAATCGACCAAC 18
|||||
DB 478 GATCGAATCGACCAAC 495

RESULT 8
US-09-385-982-497
; Sequence 497, Application US/09385982
; Patent No. 626234
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 497
; LENGTH: 587
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: (1)..(587)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-497

Query Match
71.0%; Score 14.2; DB 4; Length 587;
Best Local Similarity 84.2%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCGAATCGACCAAC 19
|||||
DB 505 GATCGAATCGACCAAC 523

RESULT 9
US-09-098-219B-1
; Sequence 1, Application US/09098219B
; Patent No. 6441277
```

```

; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard
; APPLICANT: Chelkh, No. 6441277dine
; APPLICANT: Kishore, Ganesh
; TITLE OF INVENTION: Expression of Fructose 1,6 Biphosphate
; TITLE OF INVENTION: Aldolase in Transgenic Plants
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,219B
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/049,995
; FILING DATE: 17-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: MOBT:086
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-787-1440
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1080 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-098-219B-1

Query Match
71.0%; Score 14.2; DB 4; Length 1080;
Best Local Similarity 84.2%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATCGAATCGACCAAC 19
|||||
DB 706 GTTGTGATCGACCAATCC 724

RESULT 10
US-08-976-259-22
; Sequence 22, Application US/08976259
; Patent No. 6316609
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gail H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,259
FILING DATE: Heterolith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488-0740002/EKS/CBM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 2916 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-976-259-22

```

```

Query Match
Best Local Similarity 84.2%; Score 14.2; DB 4; Length 2916;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 ATCTGAATCGACCAACC 20
      ||||| ||||| |||||
DB 189 ATCTGACCGCACCAACTG 207

```

```

RESULT 11
US-08-176-620A-5
; Sequence 5, Application US/08176620A
; Patent No. 5595904
;
GENERAL INFORMATION:
APPLICANT: Boulton, Teri G.
APPLICANT: Cobb, Melanie H.
APPLICANT: Yancopoulos, George D.
APPLICANT: Nye, Steven
APPLICANT: Panayotatos, Nikos
TITLE OF INVENTION: A Family of Map2 Protein Kinases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,620A
FILING DATE: 03-JAN-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-123
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3671 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein

```

```

FEATURE:
NAME/KEY: CDS
LOCATION: 303..2018
US-08-176-620A-5

```

```

Query Match
Best Local Similarity 84.2%; Score 14.2; DB 1; Length 3671;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 GATCTGAATCGACCAACC 19
      ||||| ||||| |||||
DB 843 GATCTCAAGCGGCGCAACC 861

```

```

RESULT 12
US-08-463-862-5
; Sequence 5, Application US/08463862
; Patent No. 5776751
;
GENERAL INFORMATION:
APPLICANT: Boulton, Teri G. et al.
TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSER: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,862
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/701,544
FILING DATE: 16-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-049
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3671 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 303..2018
US-08-463-862-5

```

```

Query Match
Best Local Similarity 84.2%; Score 14.2; DB 1; Length 3671;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 GATCTGAATCGACCAACC 19
      ||||| ||||| |||||
DB 843 GATCTCAAGCGGCGCAACC 861

```

```

RESULT 13
US-08-461-985-5
; Sequence 5, Application US/08461985

```

; Patent No. 5872006  
; GENERAL INFORMATION:  
; APPLICANT: Boulton, Teri G.  
; APPLICANT: Cobb, Melanie H.  
; APPLICANT: Yancopoulos, George D.  
; APPLICANT: Nye, Steven  
; APPLICANT: Panayiotatos, Nikos  
; TITLE OF INVENTION: A Family of Map2 Protein Kinases  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/461,985  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 800  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/176,620  
; FILING DATE: 03-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 6526-123  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3671 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 303..2018  
; US-08-461-985-5  
  
Query Match 71.0%; Score 14.2; DB 2; Length 3671;  
Best Local Similarity 84.2%; Pred. No. 1.3e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 GATCGATTCGACCAACC 19  
DB 843 GATCTCAAGCCGGCCCAACC 861  
  
RESULT 14  
US-08-458-887-5  
; Sequence 5, Application US/08458887  
; Patent No. 5914261  
; GENERAL INFORMATION:  
; APPLICANT: Boulton, Teri G. et al.  
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/458,887  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 536  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 07/701,544  
; FILING DATE: 16-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 6526-049  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3671 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 303..2018  
; US-08-458-887-5  
  
Query Match 71.0%; Score 14.2; DB 2; Length 3671;  
Best Local Similarity 84.2%; Pred. No. 1.3e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 GATCGATTCGACCAACC 19  
DB 843 GATCTCAAGCCGGCCCAACC 861  
  
RESULT 15  
US-08-390-888A-2/c  
; Sequence 2, Application US/08390888A  
; Patent No. 5916754  
; GENERAL INFORMATION:  
; APPLICANT: Nichol, Stuart T.  
; APPLICANT: Morzunov, Sergey  
; APPLICANT: Ksiazek, Thomas G.  
; APPLICANT: Rollin, Pierre E.  
; APPLICANT: Sotiropoulos, Christina F.  
; TITLE OF INVENTION: THE BAYOU HANTAVIRUS AND RELATED METHODS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
; STREET: 127 Peachtree Street, N.E., Suite 1200  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/390,888A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spratt, Gwendolyn D.  
; REGISTRATION NUMBER: 36,016  
; REFERENCE/DOCKET NUMBER: 1414.623  
; TELECOMMUNICATION INFORMATION:



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

OK nucleic - nucleic search, using sw model

Run on: May 12, 2003, 02:35:34 ; Search time 18.7143 Seconds  
(without alignments)  
1328.082 Million cell updates/sec

Title: US-09-980-277-2

Perfect score: 20

Sequence: 1 gatcgaatccgacacacg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/PCNUS\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCNUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	79.0	530	10 US-09-864-761-13599	Sequence 13599, A
2	15.8	79.0	1005	10 US-09-815-242-9151	Sequence 9151, Ap
3	15.2	76.0	457	9 US-09-918-995-2350	Sequence 2350, Ap
4	15.2	76.0	2460	9 US-09-895-288-43	Sequence 43, Appl
5	15.2	76.0	7914	12 US-10-095-718-3	Sequence 1, Appl
6	15.2	76.0	7944	12 US-10-095-718-1	Sequence 3, Appl
7	15.2	76.0	19929	9 US-09-764-891-9967	Sequence 9967, Ap
8	15.2	76.0	20907	9 US-09-764-891-9966	Sequence 9966, Ap
9	15.2	76.0	536165	9 US-09-339-964-1	Sequence 3071, Ap
10	14.8	74.0	437	10 US-09-867-701-3071	Sequence 794, App
11	14.8	74.0	479	9 US-09-918-995-794	Sequence 32024, A
12	14.8	74.0	560	10 US-09-864-761-32024	Sequence 1568, Ap
13	14.8	74.0	732	9 US-09-938-842A-1568	Sequence 15510, A
14	14.8	74.0	1985	10 US-09-864-761-15510	Sequence 455, App
15	14.4	72.0	459	10 US-09-770-444-455	Sequence 2462, Ap
16	14.4	72.0	773	10 US-09-974-300-2462	Sequence 6740, Ap
17	14.2	71.0	241	10 US-09-878-574-6740	Sequence 8698, Ap
18	14.2	71.0	421	9 US-09-918-995-8698	Sequence 3916, Ap
19	14.2	71.0	454	9 US-09-918-995-3916	

C	20	14.2	71.0	466	10 US-09-864-761-14634	Sequence 14634, A
	21	14.2	71.0	485	10 US-09-917-800A-585	Sequence 585, App
	22	14.2	71.0	572	10 US-09-974-300-1878	Sequence 1878, Ap
C	23	14.2	71.0	615	10 US-09-974-300-4348	Sequence 4348, Ap
	24	14.2	71.0	854	10 US-09-925-300-382	Sequence 382, App
C	25	14.2	71.0	1066	10 US-09-770-444-227	Sequence 227, App
	26	14.2	71.0	1080	10 US-09-923-109-1	Sequence 917, Appl
C	27	14.2	71.0	1275	10 US-09-938-842-2894	Sequence 2894, Ap
	28	14.2	71.0	1372	9 US-09-898-937-1	Sequence 1, Appl
C	29	14.2	71.0	1407	10 US-09-867-550-1891	Sequence 1891, Ap
	30	14.2	71.0	2179	10 US-10-260-046-29	Sequence 458, App
	31	14.2	71.0	2318	9 US-10-044-090-312	Sequence 22, Appl
	32	14.2	71.0	2604	9 US-09-956-004-22	Sequence 353, App
	33	14.2	71.0	2632	12 US-10-044-090-312	Sequence 5, Appl
	34	14.2	71.0	2916	10 US-09-923-109-5	Sequence 6, Appl
	35	14.2	71.0	5510	12 US-09-823-109-6	Sequence 1734, Ap
	36	14.2	71.0	10846	10 US-09-764-869-1734	Sequence 3, Appl
	37	14.2	71.0	10900	10 US-09-764-869-1734	Sequence 4, Appl
	38	14.2	71.0	21129	9 US-09-923-109-6	Sequence 14201, A
	39	14.2	71.0	21129	10 US-09-923-109-6	Sequence 4684, A
	40	14.2	71.0	147309	10 US-09-742-312-3	
	41	14.2	71.0	684973	10 US-09-263-959-1	
	42	14	70.0	657	10 US-09-893-121-4	
C	43	13.8	69.0	197	10 US-09-883-965-4349	
	44	13.8	69.0	231	10 US-09-960-352-11201	
	45	13.8	69.0	271	10 US-09-923-876-4684	

## ALIGNMENTS

RESULT 1  
US-09-864-761-13599  
Sequence 13599, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Chan, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL  
FILE REFERENCE: Aecmca-x-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13599
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023314.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
US-09-864-761-13599
```

```
Query Match
Best Local Similarity 79.0%; Score 15.8; DB 10; Length 530;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 2 ATCGAATCCGACCAACG 20
Db 2 ATCGAATCCGACCAACG 20
```

```

RESULT 2
US-09-815-242-9151/c
; Sequence 9151, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9151
; LENGTH: 1005
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1005)
US-09-815-242-9151
```

```
Query Match
Best Local Similarity 79.0%; Score 15.8; DB 10; Length 1005;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 ATCGAATCCGACCAACG 20
Db 596 ATCGAATCCGACCAACG 578
```

```

RESULT 3
US-09-918-995-2350
; Sequence 2350, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hysed, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2350
; LENGTH: 457
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(457)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2350
```

```
Query Match
Best Local Similarity 76.0%; Score 15.2; DB 9; Length 457;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 GATCGAATCCGACCAACG 20
Db 142 GATCGAATCCGACCAACG 161
```

```

RESULT 4
US-09-895-298-43
; Sequence 43, Application US/09895298
; Publication No. US20030078405A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 47 Human Secreted Proteins
; FILE REFERENCE: P2035P1
; CURRENT APPLICATION NUMBER: US/09/895,298
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/591,16
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: PCT/US99/29950
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: 60/113,006
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 60/112,809
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 2460
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-895-298-43
```

```
Query Match
Best Local Similarity 76.0%; Score 15.2; DB 9; Length 2460;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 GATCGAATCCGACCAACG 20
Db 142 GATCGAATCCGACCAACG 161
```



Best Local Similarity 85.0%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GATCGAATCCGACCAACG 20  
Db 18962 GACTGATCTGACCAACTG 18981

RESULT 9  
US-09-939-964-1  
; Sequence 1, Application US/09939964  
; Publication No. US20030054522A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosenthal, Andre  
; APPLICANT: Freiberg, Christoph  
; APPLICANT: Perret, Xavier Philippe  
; APPLICANT: Broughton, William John  
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic  
; TITLE OF INVENTION: Plasmid  
; FILE REFERENCE: CAP0068  
; CURRENT APPLICATION NUMBER: US/09/939,964  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 09/214,808  
; PRIOR FILING DATE: 1999-06-22  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: Patent Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 536165  
; TYPE: DNA  
; ORGANISM: Rhizobium  
US-09-939-964-1

Query Match 76.0%; Score 15.2; DB 9; Length 536165;  
Best Local Similarity 85.0%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GATCGAATCCGACCAACG 20  
Db 222412 GATGAGATCCGACCAACG 222431

RESULT 10  
US-09-867-701-3071  
; Sequence 3071, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Aglate, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3071  
; LENGTH: 437  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-701-3071

Query Match 74.0%; Score 14.8; DB 10; Length 437;  
Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCGAATCCGACCAAC 18  
Db 309 GATCGATCTGACCAAC 326

RESULT 11  
US-09-918-995-794

; Sequence 794, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 794  
; LENGTH: 479  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(479)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-794

Query Match 74.0%; Score 14.8; DB 9; Length 479;  
Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 ATCTGATCCGACCAACC 19  
Db 434 ACCTGATCCGATCAACC 451

RESULT 12  
US-09-864-761-32024/C  
; Sequence 32024, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL I  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/652,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32024
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF127577.1
; OTHER INFORMATION: EXPRESSED IN B1474, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: NT HIT: A1163207.2, EVALUATE 0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: BE082541.1, EVALUATE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P48552, EVALUATE 1.00e-106
US-09-864-761-32024
```

```
Query Match          74.0%; Score 14.8; DB 10; Length 560;
Best Local Similarity 88.9%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 ATCTGATTCGACCAACC 19
Db      229 ATCTGATTCGACCAACC 212
```

```

RESULT 13
; Sequence 1568, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1568
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1568
```

```
Query Match          74.0%; Score 14.8; DB 9; Length 732;
Best Local Similarity 88.9%; Pred. No. 3.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 ATCTGATTCGACCAACC 19
Db      504 ATCTGATTCGACCAACC 487

RESULT 14
```

```

US-09-864-761-15510/c
; Sequence 15510, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: neonica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15510
; LENGTH: 1985
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF127577.1
; OTHER INFORMATION: EXPRESSED IN B1474, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
US-09-864-761-15510
```

```
Query Match          74.0%; Score 14.8; DB 10; Length 1985;
Best Local Similarity 88.9%; Pred. No. 3.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 ATCTGATTCGACCAACC 19
Db      600 ATCTGATTCGACCAACC 583
```

```

RESULT 15
US-09-770-444-455
; Sequence 455, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurdan, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 455
; LENGTH: 459
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-444-455

```

```

Query Match 72.08; Score 14.4; DB 10; Length 459;
Best Local Similarity 93.88; Pred. No. 5.4e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ATCTGAATCGACAA 17
||||| |||||
Db 56 ATCTGAATCGACAA 71

```

Search completed: May 12, 2003, 04:49:50  
 Job time : 130.714 secs

GenCore version 5.1.4.P5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:04:28 ; Search time 244.143 seconds

(without alignments)  
1326.722 Million cell updates/sec

Title: US-09-980-277-2

Perfect score: 20  
Sequence: 1 gatcgaatccgacacacgcg 20

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*
  
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	92.0	687	17	BH210086
2	16.8	84.0	838	17	BH526706
3	16.4	82.0	528	17	TA11G10P
4	16.4	82.0	640	17	AO638867
5	16.4	82.0	696	12	BF621944
6	16	80.0	647	17	BH516307

7	16	80.0	745	17	BH077414
8	15	80.0	909	17	BH130484
9	15.8	79.0	184	13	EM024534
10	15.8	79.0	234	10	BB386677
11	15.8	79.0	256	10	BB357524
12	15.8	79.0	285	14	BQ236459
13	15.8	79.0	300	17	BH224590
14	15.8	79.0	302	13	BM443097
15	15.8	79.0	328	10	BB297513
16	15.8	79.0	360	9	AA462699
17	15.8	79.0	362	17	AA921320
18	15.8	79.0	390	9	AA544509
19	15.8	79.0	406	17	BH237038
20	15.8	79.0	436	17	BH224579
21	15.8	79.0	454	17	BH225175
22	15.8	79.0	458	10	BB745434
23	15.8	79.0	533	10	BE446503
24	15.8	79.0	533	12	BG739245
25	15.8	79.0	554	13	BI981975
26	15.8	79.0	557	13	BM103560
27	15.8	79.0	560	13	BJ221092
28	15.8	79.0	560	13	BI350830
29	15.8	79.0	596	13	BI706725
30	15.8	79.0	600	17	BH236984
31	15.8	79.0	606	10	AY835595
32	15.8	79.0	620	10	BY468527
33	15.8	79.0	621	10	BE189862
34	15.8	79.0	626	17	BH743324
35	15.8	79.0	628	9	AL797621
36	15.8	79.0	631	9	AL680077
37	15.8	79.0	637	17	BH449707
38	15.8	79.0	643	12	BP937817
39	15.8	79.0	643	17	BH236968
40	15.8	79.0	653	10	AA422560
41	15.8	79.0	655	9	AL655921
42	15.8	79.0	659	13	BJ071265
43	15.8	79.0	662	13	BJ064162
44	15.8	79.0	663	17	BH589281
45	15.8	79.0	667	13	BO972224

## ALIGNMENTS

RESULT 1  
LOCUS BH210086 687 bp DNA linear GSS 24-OCT-2001  
DEFINITION Sml-56U20.TF Sml Schistosoma mansoni genomic clone Sml-56U20, DNA sequence.  
ACCESSION BH210086  
VERSION BH210086.1 GI:16389299  
KEYWORDS GSS.  
SOURCE Schistosoma mansoni.  
ORGANISM Schistosoma mansoni.  
REFERENCE Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
Sneath, J., Simpson, A., Malek, J., Koo, H., Loverde, P.T. and El-Sayed, N.M.  
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sml BAC library for gene discovery and map construction  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSes: Sml-56U20.TF  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Seq primer: M13 For  
Class: BAC ends.

FEATURES  
SOURCE

Location/Qualifiers

1. 687  
/organism="Schistosoma mansoni"  
/strain="puerto rico"  
/db\_xref="taxon:6183"  
/clone="Sm1-56J20"

/note="Vector: pReloBAC11; Site\_1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pReloBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

BASE COUNT  
ORIGIN

218 a 101 c 152 g 216 t

## Query Match

Best Local Similarity 92.0%; Score 18.4; DB 17; Length 687;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCTGAATCCGACCAACCG 20

Db 348 GATCTGAATCCGACCAACCG 367

## RESULT 2

LOCUS BH526706 838 bp DNA linear GSS 13-DEC-2001  
DEFINITION BONGN32TR BONG Brassica oleracea genomic clone BONGN32, DNA sequence.

ACCESSION BH526706  
VERSION BH526706.1 GI:17734791

KEYWORDS GSS.  
SOURCE Brassica oleracea.

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 838)  
AUTHORS Town,C.D., Van Aken,S., Uterback,T. and Fraser,C.M.  
TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)

COMMENT Other GSSs: BONGN32TF

Contact: Chris Town

## TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

FEATURES  
SOURCE

1. 838  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"

/clone="BONGN32"

/clone.lib="BOGN"

/note="Vector: pHO51; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"

BASE COUNT 277 a 166 c 112 g 283 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 17; Length 838;  
Best Local Similarity 90.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCTGAATCCGACCAACCG 20  
Db 543 GATCTGAATCCGACCAACCG 562

## RESULT 3

LOCUS TALL1G10P 528 bp DNA linear GSS 13-DEC-2000  
DEFINITION T. brucei sheared genomic DNA clone 11g10, forward sequence, genomic survey sequence.

ACCESSION AL451692  
VERSION AL451692.1 GI:11831715

## KEYWORDS

GSS.

## SOURCE

Trypanosoma brucei.

## ORGANISM

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

## REFERENCE

AUTHORS

## TITLE

JOURNAL

COMMENT

1 (bases 1 to 528)

Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,

Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,

Melville,S.E., Rajandream,M.A. and Barrell,B.G.

Direct Submission

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,

Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and

nh@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR),

Rockville, MD. Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUT4 10.1) was mechanically sheared

to give a tight size distribution (4 kb). The v + i method used for the library construction is

described in detail in Smith, H. and Venter, J.C. (Making small

insert libraries for whole genome shotgun sequencing projects. In

Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available

at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

Location/Qualifiers

1. 528

/organism="Trypanosoma brucei"

/strain="TREU927"

/db\_xref="taxon:5691"

/clone="11g10"

BASE COUNT 122 a 155 c 118 g 133 t

## ORIGIN

Query Match 82.0%; Score 16.4; DB 17; Length 528;  
Best Local Similarity 94.4%; Pred. No. 4.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TCTGAATCCGACCAACCG 20  
Db 208 TCTGAATCCGACCAACCG 225

RESULT 4  
LOCUS A0638867 640 bp DNA linear GSS 08-JUL-1999  
DEFINITION 927P1-8A1.TV 927P1 Trypanosoma brucei genomic clone 927P1-8A1, DNA sequence.

ACCESSION A0638867  
VERSION A0638867.1 GI:5115577

## KEYWORDS

GSS.

## SOURCE

Trypanosoma brucei.

## ORGANISM

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 640)

El-Sayed,N., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,

Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,

Fraser,C. and Adams,M.

Determination of clone end sequences from Trypanosoma brucei TREU

TITLE

JOURNAL  
COMMENT

927/4 Pl library  
Unpublished (1999)  
Other\_GSSS: 927Pl-8A1.TP  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
For clone/filter availability, please contact Sara Melville  
(sm16@olebio.cam.ac.uk). Pl end sequences search page:  
<http://www.tigr.org/tdb/mdb/tbdb/>.  
Seq primer: 17  
Class: Pl ends.

## FEATURES

Location/Qualifiers  
1..640

/organism="Trypanosoma brucei"  
/strain="TREU927/4"  
/db\_xref="taxon:5691"  
/clone="927Pl-8A1"  
/clone\_lib="927Pl"

/note="Vector: PAD10SABIT; Site.1: Bam HI; Constructed by Sara Melville, University of Cambridge, UK and Nancy Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was isolated from *Trypanosoma brucei* (stock TREU927/4) and partially digested with Sau 3A1. DNA fragments were cloned into the Bam HI site of PAD10SABIT vector (Genbank accession U09128). The average insert size is 65 kb. Coverage: approx 4.4 X the haploid non-mitochondrial genome"

BASE COUNT 157 a 181 c 155 g 147 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 17; Length 640;  
Best Local Similarity 94.4%; Pred. No. 5.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 TCTGATCGACCAACCG 20  
||||| |||||||||  
DB 379 TCTGATCGACCAACCG 396

RESULT 5  
BF621944  
LOCUS 696 bp mRNA linear EST 17-OCT-2001  
DEFINITION HY5ME0001F16f Hordeum vulgare seedling shoot EST library  
HYCDNA0001 (Cold stress) Hordeum vulgare cDNA clone HY5ME0001F16f,  
mRNA sequence.

ACCESSION BF621944

VERSION BF621944.2 GI:13079578

## KEYWORDS

SOURCE

## ORGANISM

Hordeum vulgare.  
Hordeum vulgare.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
; Triticeae; Hordeum.  
1 (bases 1 to 696)

## REFERENCE

AUTHORS

## TITLE

JOURNAL

## COMMENT

Unpublished (2001)  
On Dec 18, 2000 this sequence version replaced gi:11885678.  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: twing@clemson.edu

Total hg bases = 167  
Seq primer: AATTAACCTCACTAAGCG  
High quality sequence stop: 485.  
Location/Qualifiers  
1..696

/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HY5ME0001F16f"  
/clone\_lib="Hordeum vulgare seedling shoot EST library  
HYCDNA0001 (Cold stress)"  
/tissue\_type="Seedling shoot"  
/lab\_host="TUC121"

/note="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI;  
Seeds were surface sterilized then germinated under aseptic  
conditions in the dark at room temperature on filter paper  
with water, nystatin and cefotaxime in covered  
crystallization dishes. Five-day old seedlings were  
incubated at 50c for 2 days. Shoots were then harvested,  
total RNA was prepared, poly(A) RNA was purified, one  
primary unamplified cDNA library was made, and 600000 pfu  
were in vivo excised to give plasmid script SK(-) cDNA  
phagemids. These steps were performed in the Tj Close  
laboratory at the University of California, Riverside  
(Choi, Close, Fenton). Phagemids were plated and picked at  
the Clemson University Genomics Institute (CUGI) (Begum,  
Palmer, Fritsch, Atkins and Wing). Plasmid DNA preparations  
DNA sequencing and sequence analysis were performed at  
CUGI (Wing, Yu, Fritsch, Henry, Simmons, Oates, Rambo, Main  
) . The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
<http://www.genome.clemson.edu/projects/barley>. To order  
this clone see <http://www.genome.clemson.edu/orders> Also  
see Close TJ, Wing R, Kleinof A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30  
(<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

BASE COUNT 183 a 146 c 197 g 170 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 12; Length 696;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATCTGATCGACCAAC 18

DB 488 GATCTGATCGACCAAC 505

## RESULT 6

BH516307

LOCUS 647 bp DNA linear GSS 13-DEC-2001

DEFINITION BOGKX42RF BOGK Brassica oleracea genomic clone BOGKX42, DNA  
sequence.

ACCESSION BH516307

VERSION BH516307.1 GI:17724397

## KEYWORDS

SOURCE

## ORGANISM

Brassica oleracea.  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 647)

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other\_GSSS: BOGKX42RF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523

Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TP  
Class: sheared ends.  
Location/Qualifiers

FEATURES  
Source  
1..647  
/organism="Brassica oleracea"  
/strain="TO100DPH3"  
/db\_xref="taxon:3712"  
/clone="BOKX42"  
/clone\_lib="BOKG"  
/note="Vector: pHOS1; Site.1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"  
BASE COUNT 147 a 205 c 110 g 185 t  
ORIGIN

Query Match 80.0%; Score 16; DB 17; Length 647;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TCTGATCCGACCAC 18  
|||||  
Db 565 TCTGATCCGACCAC 580

RESULT 7  
BH077414  
LOCUS BH077414 745 bp DNA linear GSS 18-JUN-2001  
DEFINITION RPT-24-377116.TV RPT-24 Mus musculus genomic clone RPT-24-377116  
, DNA sequence.  
ACCESSION BH077414  
VERSION BH077414.1 GI:14897011  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 745)  
Zhao, S., Mierman, W., Malek, J., Shatsman, S., Akintel, B., Levins, M.,  
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
Russell, D., de Jong, P. and Fraser, C. M.  
Mouse BAC End Sequences from Library RPT-24  
Unpublished (1999)  
Other\_GSSs: RPT-24-377116.TJ  
COMMENT  
Contact: Shanying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPT-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@tigr.org). Clones may be purchased from BACPAC  
Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end  
page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 377 row: 1 column: 16  
Seq primer: 17  
Class: BAC ends.

FEATURES  
Source  
1..745  
Location/Qualifiers

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPT-24-377116"  
/clone\_lib="RPT-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTARBA1; Site.1: BamHI; Site.2: BamHI;  
RPT-24 Mouse BAC Library produced by Pieter de Jong. The  
library was cloned in the pTARBA1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J

BASE COUNT 152 a 236 c 144 g 213 t  
ORIGIN

Query Match 80.0%; Score 16; DB 17; Length 745;  
Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CTGATCCGACCAC 19  
|||||  
Db 730 CTGATCCGACCAC 745

RESULT 8  
BH130484 909 bp DNA linear GSS 07-AUG-2001  
LOCUS BH130484  
DEFINITION ENTOL69TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
genomic, DNA sequence.  
ACCESSION BH130484  
VERSION BH130484.1 GI:15088953  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Eukaryota; Entamoebidae; Entamoeba.  
Eukaryota; Entamoebidae; Entamoeba.  
1 (bases 1 to 909)  
Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
Determination of clone end sequences from Entamoeba histolytica  
BHI:IMSS sheared DNA library (2001)  
Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjl@tigr.org  
Clones are derived from the Entamoeba histolytica BHI:IMSS sheared  
DNA library  
Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 48  
High quality sequence stop: 770.  
Location/Qualifiers

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
Source  
1..909  
/organism="Entamoeba histolytica"  
/strain="BHI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOS1; Site.1: Bst I; Constructed at the  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.) The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Bartell, Oxford University Press, 1999)."  
Bartell, Oxford University Press, 1999)."

BASE COUNT 432 a 91 c 142 g 244 t  
ORIGIN

Query Match 80.0%; Score 16; DB 17; Length 909;  
Best Local Similarity 100.0%; Pred. No. 9.3e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATCTGATCCGACCA 17  
|||||  
Db 33 ATCTGATCCGACCA 48

[illegible]

VERSION	BBS98677.1	GI:9218073
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 234)	
AUTHORS	Kono H., Aizawa K., Akahira S., Akiyama J., Arakawa T., Carninci P., Endo T., Fukuda S., Fukushima Y., Hara A., Hayatsu N., Hirotsue T., Horii F., Ishii T., Ishikawa J., Ishikawa T., Itoh M., Izawa M., Kadoya K., Kagawa T., Kai C., Kawai J., Kikuchi N., Kiyoawa H., Koijima Y., Kondo S., Koya S., Kuwahara C., Kusakabe M., Matsuyama T., Miki R., Mizuno Y., Nakamura M., Oda H., Okazaki Y., Ono T., Owa C., Satto H., Sakai C., Sato K., Shibata Y., Shigemoto Y., Shinagawa A., Shiraki T., Sogabe Y., Sugihara Y., Suzuki H., Suzuki H., Tagawa A., Takahashi F., Tomimaga N., Toya T., Tsunoda Y., Watanuki A., Watanabe S., Yamamori T., Yamataka I., Yano R., Yasunishi A., Yokota T., Yoshida K., Yoshiki A., Yoshino M., Muramatsu M. and Hayashizaki Y.	
TITLE	RIKEN Mouse ESTs (Kono H., et al.)	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: yoshinobe hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Sehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp, url:http://genome.gsc.riken.go.jp/ Carninci P., Nishiyama Y., Westover A., Itoh M., Nagaoka S., Sasaki N., Okazaki Y., Muramatsu M. and Hayashizaki Y. Thermostabilization and thermoactivation of the thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (12), 520-524 (1998) Itoh M., Katsuna T., Akiyama J., Shibata K., Izawa M., Kawai J., Tomaru Y., Carninci P., Shibata Y., Ozawa Y., Muramatsu M., Okazaki Y. and Hayashizaki Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci P. and Hayashizaki Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site ( <a href="http://genome.rtc.riken.go.jp">http://genome.rtc.riken.go.jp</a> ) for further details. Location/Qualifiers 1..234 /organism="Mus musculus" /strain="G57BL/6J" /db_xref="taxon:10090" /clone="C330009119" /clone_1lb="RIKEN full-length enriched, ES cells" /cell_type="ES cells" /lab_host="SOLR" note=Site.1: XhoI; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAAGAGAGAAGATCCAGACGCCTTTTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAAGAGAGATTCGCAATTAAATTAATTCACCCCCCCCCCCC 3']"	
BASE COUNT	76 a	59 c     29 g     70 t
Query Match	79.0%; Score 15.8; DB 10;	Length 234;
Best Local Similarity	89.5%; Pred. No. 7.4e+02;	



Best Local Similarity 89.5%; Pred. No. 8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCGAATCGACCAACC 19  
|||||  
Db 204 GATCGAATCGACCAACC 222

## RESULT 13

LOCUS BH224590 300 bp DNA linear GSS 08-NOV-2001  
DEFINITION 1006119H11.2EL.Y2 1006 - Rescemu Grid G Zea mays genomic, DNA  
sequence.

ACCESSION BH224590  
VERSION BH224590.1 GI:16821698  
KEYWORDS GSS.  
SOURCE Zea mays.  
ORGANISM Zea mays.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 300)

REFERENCE Althors  
TITLE Maize genomic sequences found using engineered Rescemu transposon  
JOURNAL Unpublished (2001)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Possible ligation site of ends cut by 2 different endonucleases.  
Reverse complemented post-ligation sequence from source sequence.  
Plate: 1006119 row: 44  
Class: transposon-tagged.

FEATURES  
SOURCE Location/Qualifiers  
1..300

/organism="Zea mays"  
/cultivar="mixed background W23/A188/B73"  
/db\_xref="taxon:4577"  
/clone\_lib="1006 - Rescemu Grid G"  
/tissue\_type="leaf"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: leaf; Vector: Rescemu (engineered from  
pBlueScript backbone); Site\_1: BamHI; Site\_2: BglII;  
Rescemu is a 4.9 kb, modified maize Mu transposon  
designed to allow plasmid rescue from total genomic DNA.  
Mu elements insert preferentially into transcription  
units. For more information on Rescemu, go to the web  
site 'www.zmbl.jastat.edu' and follow the links for  
'Rescemu'. Grid G was grown at Stanford in 2000. DNA was  
extracted from leaf punches, double digested using BamHI  
and BglII, and ligated to form circular plasmids. DH10B  
cells were transformed and then screened on LB plates with  
ampicillin."

BASE COUNT 63 a 64 c 81 g 91 t 1 others  
ORIGIN

Query Match 79.0%; Score 15.8; DB 17; Length 300;  
Best Local Similarity 89.5%; Pred. No. 8.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCGAATCGACCAACC 19  
|||||  
Db 73 GATCGAATCGACCAACC 91

RESULT 14  
BM443097/c 302 bp mRNA linear EST 23-JUL-2002  
LOCUS BM443097  
DEFINITION EBR002\_SQ002\_118\_R root, 3 week, hydroponic grown, low nitrogen, cv

ACCESSION BM443097  
VERSION BM443097.2 GI:21947759  
KEYWORDS EST.  
SOURCE Hordeum vulgare.  
ORGANISM Hordeum vulgare.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 302)  
AUTHORS Hedley, P., Liu, H., Caldwell, D., McCallum, N., Madie, S., Cardie, L.,  
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.  
TITLE Development of Barley Transcriptome Resources  
JOURNAL Unpublished (2001)  
COMMENT On Feb 1, 2002 this sequence version replaced gi:18473872.  
Contact: Waugh R, Marshall DF  
Genome Dynamics/Computational Biology  
Scottish Crop Research Institute  
Invergowrie, Dundee, DD2 5DA, Scotland, UK  
Tel: 00 44 1382 562731  
Fax: 00 44 1382 562426  
Email: est@scri.sari.ac.uk

FEATURES  
SOURCE All sequence has a phred quality score of 20 or over  
Seq primer: M13 reverse  
Location/Qualifiers  
1..302

/organism="Hordeum vulgare"  
/cultivar="Optic"  
/db\_xref="taxon:4513"  
/clone\_lib="EBR002\_SQ002\_118"  
/clone\_lib="root, 3 week, hydroponic grown, low nitrogen,  
cv Optic, EBR002"  
/tissue\_type="root"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/note="Vector: pSPOR1; Site\_1: Sal I; Site\_2: Not I;  
Non-normalised library, directionally cloned into pSPOR1.  
Derived from roots of 3 week old Nitrogen stressed barley  
plants. Developed as part of the barley transcriptome  
resources of BSRG/SEERD funded cereal IGF (Investigating  
Gene Function) project."

BASE COUNT 54 a 72 c 92 g 84 t  
ORIGIN

Query Match 79.0%; Score 15.8; DB 13; Length 302;  
Best Local Similarity 88.5%; Pred. No. 8.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATCGAATCGACCAACC 19  
|||||  
Db 253 GATCGAATCGACCAACC 235

## RESULT 15

LOCUS BB297513 328 bp mRNA linear EST 10-JUN-2000  
DEFINITION BB297513 RIKEN full-length enriched, 9.5 days embryo parthenogenote  
Mus musculus cDNA clone B130055M05 3', mRNA sequence.

ACCESSION BB297513  
VERSION BB297513.1 GI:8998218  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 328)  
AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, T., Arikawa, T., Carninci,  
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
Katsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,



GenCore version 5.1.4 p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic acid search, using sw model

Run on: May 12, 2003, 01:02:58 ; Search time 99 Seconds  
(without alignments)  
5879.385 Million cell updates/sec

Title: US-09-980-277-3

Perfect score: 20

Sequence: 1 atatacgcgcacgcctc 20

Scoring table: IDENTITY NJC  
Gapop 10.0 , Gapept 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pln:\*  
35: em\_hlg\_rod:\*  
36: em\_hlg\_mem:\*  
37: em\_hlg\_vtl:\*  
38: em\_sy:\*  
39: em\_higo\_hum:\*  
40: em\_higo\_mus:\*  
41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20	100.0	121	3	SCMRPTAA
C 2	17.4	87.0	146370	2	AC094913
C 3	16.8	84.0	1356	8	AF010508
C 4	16.8	84.0	106685	8	AC023754
C 5	16.8	84.0	106760	9	HDAC002990
C 6	16.8	84.0	119535	2	AC129382
C 7	16.8	84.0	168949	9	AC046134
C 8	16.8	84.0	184321	2	AC025890
C 9	16.4	82.0	3226	9	AC027375
C 10	16.4	82.0	4695	3	AF387518
C 11	16.4	82.0	5878	3	AF121666
C 12	16.4	82.0	14086	1	AE009916
C 13	16.4	82.0	40123	2	AC014130
C 14	16.4	82.0	65961	2	AC016523
C 15	16.4	82.0	80866	2	AC012649
C 16	16.4	82.0	109155	8	ATT20K12
C 17	16.4	82.0	112314	9	AC004002
C 18	16.4	82.0	154411	3	AC073909
C 19	16.4	82.0	163307	3	AC011703
C 20	16.4	82.0	167688	3	AC011702
C 21	16.4	82.0	173281	3	AC012161
C 22	16.4	82.0	300994	3	AE003506
C 23	16.4	82.0	305124	2	AC073858
C 24	16	80.0	691	3	DFCH3C1
C 25	16	80.0	155934	2	AC116613
C 26	16	80.0	220962	2	AC126272
C 27	15.8	79.0	678	9	HSB340686
C 28	15.8	79.0	1887	9	AF064769
C 29	15.8	79.0	4944	3	DMRNOT2
C 30	15.8	79.0	5482	10	RNRG9
C 31	15.8	79.0	6075	6	AX346453
C 32	15.8	79.0	6207	6	AX339169
C 33	15.8	79.0	6922	6	AX251287
C 34	15.8	79.0	6922	6	AX252032
C 35	15.8	79.0	6922	6	AX344406
C 36	15.8	79.0	6922	6	AX348811
C 37	15.8	79.0	7631	6	AX345762
C 38	15.8	79.0	8076	6	AX348340
C 39	15.8	79.0	9543	6	AX344641
C 40	15.8	79.0	11844	9	AL391691
C 41	15.8	79.0	11845	1	AE011176
C 42	15.8	79.0	14344	5	AF124048
C 43	15.8	79.0	14816	10	MM0276492
C 44	15.8	79.0	16522	2	AC013170
C 45	15.8	79.0	19459	6	AX344209

## ALIGNMENTS

RESULT 1  
SCMRPTAA/C 121 bp DNA linear INV 26-APR-1993  
LOCUS S.mansoni tandem repeat units.  
DEFINITION M61098  
ACCESSION M61098  
VERSION M61098.1 GI:161064  
KEYWORDS  
SOURCE  
ORGANISM  
S.mansoni (strain Egyptian) DNA, clone psm1-7.  
Schistosoma mansoni  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE  
1 (bases 1 to 121)  
Hamburger,J., Turetski,T., Kapeller,I. and Deresiewicz,R.  
TITLE  
Highly repeated short DNA sequences in the genome of Schistosoma mansoni recognized by a species-specific probe

JOURNAL Mol. Biochem. Parasitol. 44 (1), 73-80 (1991)  
 MEDLINE 91187067  
 PUBMED 2011155  
 FEATURES Location/Qualifiers  
 SOURCE 1. 121  
 /organism="Schistosoma mansoni"  
 /db\_xref="taxon:6183"  
 BASE COUNT 36 a 20 c 26 g 39 t  
 ORIGIN  
 Query Match 100.0%; Score 20; DB 3; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 9.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATATTACGCCACGCTCTC 20  
 ||||||||||||||||  
 Db 110 ATATTACGCCACGCTCTC 91  
 RESULT 2  
 AC094913/c 146970 bp DNA linear HMG 10-JUL-2002  
 LOCUS Rattus norvegicus clone CH230-6C20, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
 59 unordered pieces.  
 AC094913  
 AC094913.3 GI:21716171  
 HMG; HMG\_PHASEL.  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 146970)  
 Kuzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbieri,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brieva,M., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,J., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,  
 Homsi,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Melker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G.,  
 Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,D., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rivers,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Savary,G.,  
 Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,  
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,K., Tang,H.,  
 Sutton,A., Sytek,A., Tabor,P., Tameris,A., Tameris,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Umami,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,O.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczyk,R., Woodson,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstock,G. and Gibbs,R.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 TITLE  
 AUTHORS  
 REFERENCE  
 JOURNAL  
 TITLE  
 AUTHORS  
 COMMENT  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 146970)  
 Morley K.C.  
 Direct Submission  
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 146970)  
 Morley K.C.  
 Direct Submission  
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 9, 2002 this sequence version replaced gi:17941710.  
 ----- Genome Center -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information -----  
 Center project name: GBRX  
 Center clone name: CH230-6C20  
 ----- Summary Statistics -----  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 92271 bases at least Q40  
 Consensus quality: 97706 bases at least Q30  
 Consensus quality: 102314 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 59 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1179: contig of 1179 bp in length  
 \* 1180 1279: gap of unknown length  
 \* 1280 2496: contig of 1217 bp in length  
 \* 2497 2596: gap of unknown length  
 \* 2597 4171: contig of 1574 bp in length  
 \* 4171 4270: gap of unknown length  
 \* 4271 5575: contig of 1305 bp in length  
 \* 5576 5675: gap of unknown length  
 \* 5676 7019: contig of 1344 bp in length  
 \* 7020 7119: gap of unknown length  
 \* 7120 8691: contig of 1572 bp in length  
 \* 8692 8791: gap of unknown length  
 \* 8792 10183: contig of 1392 bp in length  
 \* 10184 10283: gap of unknown length  
 \* 10284 11471: contig of 1188 bp in length  
 \* 11472 11571: gap of unknown length  
 \* 11572 12723: contig of 1152 bp in length  
 \* 12724 12823: gap of unknown length  
 \* 12824 14403: contig of 1580 bp in length  
 \* 14404 14503: gap of unknown length  
 \* 14504 15527: contig of 1024 bp in length  
 \* 15528 15627: gap of unknown length  
 \* 15628 17611: contig of 1984 bp in length  
 \* 17612 17711: gap of unknown length  
 \* 17712 19429: contig of 1718 bp in length  
 \* 19430 19529: gap of unknown length  
 \* 19530 21227: contig of 1698 bp in length  
 \* 21228 21327: gap of unknown length  
 \* 21328 22376: contig of 1049 bp in length  
 \* 22377 22477: gap of unknown length  
 \* 22478 24109: contig of 1533 bp in length  
 \* 24109: gap of unknown length

```

* 24110 26423: contig of 2314 bp in length
* 26424 26523: gap of unknown length
* 26524 27957: contig of 1434 bp in length
* 27958 28058: gap of unknown length
* 28058 29638: contig of 1581 bp in length
* 29639 29739: gap of unknown length
* 29739 31847: contig of 2108 bp in length
* 31847 31947: gap of unknown length
* 31947 34143: contig of 2196 bp in length
* 34143 34243: gap of unknown length
* 34243 36377: contig of 2134 bp in length
* 36377 36476: gap of unknown length
* 36476 38561: contig of 2085 bp in length
* 38561 38661: gap of unknown length
* 38661 39903: contig of 1242 bp in length
* 39903 40003: gap of unknown length
* 40003 41958: contig of 1955 bp in length
* 41958 42058: gap of unknown length
* 42058 43625: contig of 1567 bp in length
* 43625 43725: gap of unknown length
* 43725 45011: contig of 1286 bp in length
* 45011 45112: gap of unknown length
* 45112 47277: contig of 2166 bp in length
* 47277 47377: gap of unknown length
* 47377 49421: contig of 2044 bp in length
* 49421 49521: gap of unknown length
* 49521 51042: contig of 1521 bp in length
* 51042 51143: gap of unknown length
* 51143 53483: contig of 2340 bp in length
* 53483 53583: gap of unknown length
* 53583 55075: contig of 1493 bp in length
* 55075 55175: gap of unknown length
* 55175 57330: contig of 2155 bp in length
* 57330 57430: gap of unknown length
* 57430 60307: contig of 2877 bp in length
* 60307 60407: gap of unknown length
* 60407 62705: contig of 2298 bp in length
* 62705 62805: gap of unknown length
* 62805 64827: contig of 2022 bp in length
* 64827 64928: gap of unknown length
* 64928 67634: contig of 2707 bp in length
* 67634 67735: gap of unknown length
* 67735 69904: contig of 2169 bp in length
* 69904 70004: gap of unknown length
* 70004 72964: contig of 2961 bp in length
* 72964 73064: gap of unknown length
* 73064 75860: contig of 2796 bp in length
* 75860 75960: gap of unknown length
* 75960 77903: contig of 1943 bp in length
* 77903 78003: gap of unknown length
* 78003 80743: contig of 2740 bp in length
* 80743 80843: gap of unknown length
* 80843 83862: contig of 3019 bp in length
* 83862 83962: gap of unknown length
* 83962 86802: contig of 2840 bp in length
* 86802 86902: gap of unknown length
* 86902 90056: contig of 3154 bp in length
* 90056 90156: gap of unknown length
* 90156 92842: contig of 2685 bp in length
* 92842 92942: gap of unknown length
* 92942 94466: contig of 1525 bp in length
* 94466 94566: gap of unknown length
* 94566 97310: contig of 2744 bp in length
* 97310 97410: gap of unknown length
* 97410 99793: contig of 2383 bp in length
* 99793 99893: gap of unknown length
* 99893 99894: gap of unknown length
* 99894 103484: contig of 3591 bp in length
* 103484 103584: gap of unknown length
* 103584 106611: contig of 3027 bp in length
* 106611 106711: gap of unknown length
* 106711 110006: contig of 3295 bp in length
* 110006 110106: gap of unknown length
* 110106 113951: contig of 3845 bp in length

```

```

* 113952 114051: gap of unknown length
* 114052 118104: contig of 4053 bp in length

Query Match
Best Local Similarity 87.0%; Score 17.4; DB 2; Length 146970;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TATTAACGCCACGCTCTC 20
Db 139691 TATTAACGCCACGCTTTC 139673

RESULT 3
AY010508/c 1356 bp DNA linear INV 15-MAR-2002
LOCUS Scallioia sp. tRNA-Thr, tRNA-Gly, and tRNA-Val genes, complete
DEFINITION sequence, and 16S ribosomal RNA gene, partial sequence;
mitochondrial genes for mitochondrial products.
ACCESSION AY010508
VERSION AY010508.1 GI:16554857
KEYWORDS
SOURCE
ORGANISM Scallioia sp.
Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
NeotaenioGLOSSA; Cerithioidea; Scallioidea; Scallioia.
REFERENCE 1 (bases 1 to 1356)
AUTHORS Lydeard,C., Holznagel,W.E., Glaubrecht,M. and Ponder,W.F.
Molecular phylogeny of a circum-global, diverse gastropod
superfamily (Cerithioidea: Mollusca: Caenogastropoda): pushing the
deepest phylogenetic limits of mitochondrial LSU rDNA sequences
Mol. Phylogenet. Evol. 22 (3), 399-406 (2002)
JOURNAL
MEDLINE 21861977
PUBMED 11864164
REFERENCE 2 (bases 1 to 1356)
AUTHORS Lydeard,C., Holznagel,W.E., Glaubrecht,M. and Ponder,W.F.
Direct Submission
TITLE Submitted (17-OCT-2000) Biodiversity and Systematics, Department of
JOURNAL Biological Sciences, The University of Alabama, Box 870345,
Tuscaloosa, AL 35487-0345, USA
FEATURES
source
1. 1356
/organism="Scallioia sp."
/organelle="mitochondrion"
/db_xref="taxon:145863"
1. 64
/product="tRNA-Thr"
trna /product="tRNA-Thr"
65. 128
trna /product="tRNA-Gly"
129. 195
trna /product="tRNA-Val"
196. >1356
/product="16S ribosomal RNA"
BASE COUNT 517 a 188 c 218 g 433 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 3; Length 1356;
Best Local Similarity 90.0%; Pred. No. 4.5e+07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTAACGCCACGCTCTC 20
Db 1125 ATATTAACGCCACGCTTTC 1106

RESULT 4
AC023754 100685 bp DNA linear PLN 13-SEP-2000
LOCUS Arabidopsis thaliana chromosome I BAC F1B16 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC023754
VERSION AC023754.3 GI:7212003
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.

```

ORGANISM	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE	1 (bases 1 to 100685)
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altali,H., Nguyen,M., Lam,B., Southwick,A., Bel,O., Buehler,E., Buehler,E., Chao,Q., Chin,C., Chio,J., Choi,E., Gonzalez,A., Hong,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M., Lenz,C., Liu,A., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Shim,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 100685)
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altali,H., Nguyen,M., Lam,B., Southwick,A., Bel,O., Buehler,E., Chin,C., Chio,J., Choi,E., Dunn,P., Gonzalez,A., Hong,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shin,P., Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
JOURNAL	Direct Submission
TITLE	Submitted (17-FEB-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE	3 (bases 1 to 100685)
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altali,H., Nguyen,M., Lam,B., Southwick,A., Bel,O., Buehler,E., Chin,C., Chio,J., Choi,E., Dunn,P., Gonzalez,A., Hong,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shin,P., Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
JOURNAL	Direct Submission
TITLE	Submitted (09-MAR-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE	4 (bases 1 to 100685)
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altali,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A. and Davis,R.W.
JOURNAL	Direct Submission
TITLE	Submitted (13-SEP-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT	On Mar 9, 2000 this sequence version replaced gi:7143418. Bases 1-32,040 of IGF clone F1B16 overlap with bases 97,508-129,547 of IGF clone F10A5, gb AC006434. email for correspondence: arabidopses.stanford.edu
FEATURES	Genes with similarity to proteins in the databases are named 'putative', 'like' or 'similar to'. Genes that have a similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The gene prediction programs used to predict genes include: Gail (Informatics Group), Oak Ridge National Laboratory, http://compbio.ornl.gov/section/index.html, GENSCAN (Chris Burge, http://genes.mit.edu/GENSCAN/), FEXA (Victor Solovyev, http://genomic.sanger.ac.uk/gf/gf.shtml), and MetaplanGene (S.M. Habsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).
SOURCE	Location/Qualifiers
	1..100685
	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"
	/chromosome="1"
	/clone="F1B16"
misc_feature	1..32040
	/note="overlap with bases 97,508-129,547 of IGF clone F10A5, gb AC006434. See GenBank record for BAC clone F10A5 for annotation in this region."
gene	complement(32986..33822)
	/gene="F1B16.1"
	/note="Unknown protein; unknown protein"
	/codon_start=1
	/protein_id="F1B16.1"
	/db_xref="GI:1012044"
	/translation="NALPYYVSPSSSTLSPEKSKPPPHSIRLSISPTVDRNRKLN SSFRCSSSSSEKHNAKSPKSDIVLEPLPLPLVLPFGALILPLOTIEPKYRLMHT ILQSDILFGVYVSDSVGSAEVAECVGEVYHERLVDDPEFLVCGKGRPEYVNVYT KPYLVGVMLEDRPGSEGNDSIANEVYLMKEVTRLSNRLNGAAKEVODLRROF PLPSFVFSSTFEGAPRQQLLEEDTFAARKRENETLRNTLNTLTAASVNDYPS SS"
gene	43756..46560
	/gene="F1B16.2"
	/join(43756..44347,44507..44562,44700..44966,45493..45755,46200..46560)
CDS	/gene="F1B16.2"
	/note="Similar to cytokinin oxidase"
	/codon_start=1
	/protein_id="F1B16.2"
	/db_xref="GI:10120443"
	/translation="MTSFLILTAICLLIIVGLVNGPSELLRIGALIDVGHPTVHP SDIASVSDGMLKSPPEPLAVLHPSAEIVAVLVPYAGSAFAPVAGRHGHSING OAAAGRVVYVEMNGVSTGPKLVYRDEMVTYVWGVEWVDTLKEHGLAPKSWT DYLIVYGFGLSNAGISGQAHGHPQISLVLEIDVYGGQGITFARISLEAPQRYL WRIYSSPKVPTPEDETLISMGQIKEDVYBEPYVDSGLVNNMRSSPSRPPRYI SSVSSNSVYICLETIKATYHDSDELVDQEVETLAKNLPITPTSTPTDLDYDFDR VKRELKRLKNLMEVPHWNLVFPKRSYSDKGVFGILGNKTSGLITIPKMD KMRDSALVPEDEVEYVALRSALVDGERTGKELYLDQKRLIECEQAKINAKO LYLPHAAIQEWWVAHFQDKWDFRRSKAFEPDRIILATGQRIQNPSSLSLPESSSSS AASP"
gene	complement(46848..48185)
	/gene="F1B16.3"
CDS	complement(join(46848..46943,47040..47128,47474..47537,47773..47875,47963..48066,48156..48185))
	/gene="F1B16.3"
	/codon_start=1
	/product="Putative ubiquitin-conjugating enzyme"
	/protein_id="F1B16.3"
	/db_xref="GI:10120441"
	/translation="MSSSGAPSRKTLKATLNRLQELVDMQNPPTGFKRYVDNLQ KWLIEVIGAPGLVADYVLDQVDPPEHPESPQVFLPAPLPHYSNGHCLDI LYDWSMPMTAVSTICISILMSLSSSTKRPDRTYVDRNGLSPKRETPMWFHDK V"
gene	52291..53501
	/gene="F1B16.4"
CDS	join(52291..52584,52917..53501)
	/gene="F1B16.4"
	/note="Hypothetical protein"
	/codon_start=1
	/protein_id="F1B16.4"
	/db_xref="GI:10120438"
	/translation="MEDFVRHESGSLRGTLDSRYAKAVOCIVEVDIGREVELC NITLINDLPGRRRGPRFALSRKSELSSGSEPSLENEHITIKTLISQLOVER EGYCNQLEOVISPEFLAGBSSSVYTGALDAMRHGSSLEPIQLNSYRRRT ISHDVFKLISGLQSLSDENTSSSLQGLVGGPQHMAKYTRKLPETSVAILR AMLQHFLLPVSFLLKCNNTSDSKLQWNNISLIVAKYTIKRSKQWMLKOD FRRKRYPLMLILFP"
gene	complement(54491..56307)
	/gene="F1B16.5"
CDS	complement(join(54491..55074,55197..55351,55498..55769,55872..56014,56253..56307))
	/gene="F1B16.5"
	/note="Hypothetical protein"
	/codon_start=1
	/protein_id="F1B16.5"
	/db_xref="GI:10120445"
	/translation="MGTDFVYVSTCINSYNSRGLIMELAFILRGADVYVITNQ KPLEDEVYVSLERKMDRGVVISAKQKQVDTSLKADILVLAIVAGKLVAVLE NVYVYLKILIMWIHNRGMYFNADVYKLPVPAQAMTIDSGIKNRTGARKIGL KPYVYHIGNSKILMEVADSVAKRVLNHEHRELSVGRNEDLLPQIINSVARGGDL FLAFHSLSIRIKERKLOVTPMAVAVVSGDSKQFELELNFRERKLENFVHEVY KTVLVAYLIAIDVLYVNSQARGECFRTTEAMAKFLPYLSTAGSGMELIVVGTG



AC129382  
 LOCUS 119535 bp DNA linear HTG 02-AUG-2002  
 DEFINITION Rattus norvegicus clone CH230-91018, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 56 unordered pieces.  
 AC129382  
 VERSION AC129382.2 GI:22038396  
 KEYWORDS HTG: HTGS\_PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 119535)  
 Muzny,D.M., Adams,C., Adio-Odola,B., Ali-Osman,F.R., Allen,C.,  
 Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,  
 Barabara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,  
 Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,  
 Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gumarate,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hughes,M., Holloway,C., Hollins,B.,  
 Homs,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvath,J., Kovar,C.,  
 Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,W., Lonsleged,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheswari,M., Kapur,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhin,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabhat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okunolu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Rao,Y.,  
 Rives,M., Rojas,A., Rojuben,R., Rolfe,M., Ruiz,S., Savary,G.,  
 Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,T.,  
 Sodergren,E., Sotake,T., Sparks,A., Stanley,H., Stone,K., Tang,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Umami,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G., and Gibbs,R.  
 Direct Submission  
 Title Unpublished  
 Journal 2 (bases 1 to 119535)  
 Reference Worley,K.C.  
 Author Title Direct Submission  
 Journal Submitted (29-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 119535)  
 Worley,K.C.  
 Direct Submission  
 Submitted (02-AUG-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Aug 1, 2002 this sequence version replaced gi:22001463.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
 Center project name: G1XG  
 Center clone name: CH230-91018  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 59977 bases at least Q40  
 Consensus quality: 67676 bases at least Q30  
 Consensus quality: 73666 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see <http://www.hgsc.bcm.tmc.edu/docs/genbank-draft-data.html>).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 56 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 1003: contig of 1002 bp in length  
 1103: gap of unknown length  
 1103: contig of 1006 bp in length  
 2109: gap of unknown length  
 2209: contig of 1635 bp in length  
 3843: gap of unknown length  
 3844: contig of 1217 bp in length  
 3944: gap of unknown length  
 5161: gap of unknown length  
 5260: contig of 1330 bp in length  
 5261: gap of unknown length  
 6590: contig of 1196 bp in length  
 6691: gap of unknown length  
 7886: contig of 1083 bp in length  
 7887: gap of unknown length  
 9069: contig of 1083 bp in length  
 9070: gap of unknown length  
 9169: contig of 1770 bp in length  
 10339: gap of unknown length  
 10940: contig of 1336 bp in length  
 11040: gap of unknown length  
 12437: contig of 1113 bp in length  
 12438: gap of unknown length  
 12537: contig of 1460 bp in length  
 12537: gap of unknown length  
 13650: contig of 1460 bp in length  
 13650: gap of unknown length  
 13651: contig of 1460 bp in length  
 13651: gap of unknown length  
 15173: contig of 1460 bp in length  
 15173: gap of unknown length  
 15174: contig of 1673 bp in length  
 15274: gap of unknown length  
 16947: gap of unknown length  
 16947: contig of 1597 bp in length  
 17046: gap of unknown length  
 17046: contig of 1597 bp in length  
 18643: gap of unknown length  
 18644: gap of unknown length  
 20203: contig of 1460 bp in length  
 20204: gap of unknown length  
 20303: gap of unknown length  
 21681: contig of 1378 bp in length  
 21681: gap of unknown length  
 21782: contig of 1625 bp in length  
 21782: gap of unknown length  
 23406: contig of 1395 bp in length  
 23406: gap of unknown length  
 23506: contig of 1395 bp in length  
 23506: gap of unknown length  
 24901: contig of 1117 bp in length  
 24901: gap of unknown length  
 25002: contig of 1117 bp in length  
 25002: gap of unknown length  
 26118: contig of 1296 bp in length  
 26118: gap of unknown length  
 26219: contig of 1296 bp in length  
 26219: gap of unknown length  
 27517: contig of 1437 bp in length  
 27517: gap of unknown length  
 29053: contig of 1954 bp in length  
 29053: gap of unknown length  
 29154: contig of 1954 bp in length  
 31107: gap of unknown length  
 31207: gap of unknown length  
 33384: contig of 2177 bp in length  
 33384: gap of unknown length  
 33385: gap of unknown length  
 33484: contig of 2223 bp in length  
 33484: gap of unknown length  
 35708: gap of unknown length  
 35708: contig of 2165 bp in length  
 35808: gap of unknown length  
 37973: gap of unknown length  
 38072: contig of 1032 bp in length  
 38073: gap of unknown length  
 39105: contig of 1558 bp in length  
 39205: gap of unknown length  
 40762: contig of 1558 bp in length  
 40762: gap of unknown length  
 40763: gap of unknown length  
 40862: gap of unknown length

```

* 40863 42690: contig of 1828 bp in length
* 42691 42790: gap of unknown length
* 44621 44621: contig of 1831 bp in length
* 44622 44721: gap of unknown length
* 44722 46492: contig of 1771 bp in length
* 46493 46592: gap of unknown length
* 46593 48535: contig of 1943 bp in length
* 48536 48635: gap of unknown length
* 48636 49726: contig of 1091 bp in length
* 49727 49826: gap of unknown length
* 49827 51434: contig of 1607 bp in length
* 51434 53995: gap of unknown length
* 53995 54094: gap of unknown length
* 54095 56420: contig of 2326 bp in length
* 56421 56520: gap of unknown length
* 56521 57610: contig of 1090 bp in length
* 57611 59464: contig of 1754 bp in length
* 59465 59565: gap of unknown length
* 59565 61645: contig of 2080 bp in length
* 61645 61745: gap of unknown length
* 61745 63220: contig of 1476 bp in length
* 63221 63320: gap of unknown length
* 63321 66614: contig of 3294 bp in length
* 66615 66714: gap of unknown length
* 66715 68357: contig of 1642 bp in length
* 68357 68456: gap of unknown length
* 68457 70817: contig of 2361 bp in length
* 70818 70917: gap of unknown length
* 70918 72489: contig of 1572 bp in length
* 72490 72589: gap of unknown length
* 72590 74365: contig of 1776 bp in length
* 74365 74465: gap of unknown length
* 74466 77170: contig of 2705 bp in length
* 77171 79757: contig of 2487 bp in length
* 79758 79857: gap of unknown length
* 79858 83585: contig of 3728 bp in length
* 83586 83686: gap of unknown length
* 83686 85419: contig of 1734 bp in length
* 85420 85519: gap of unknown length
* 85520 87286: contig of 1767 bp in length
* 87287 87387: gap of unknown length
* 87387 89241: contig of 1855 bp in length
* 89242 89341: gap of unknown length
* 89342 91596: contig of 2255 bp in length
* 91597 91696: gap of unknown length
* 91697 95810: contig of 4114 bp in length
* 95811 95910: gap of unknown length
* 95911 99718: contig of 3808 bp in length
* 99719 99818: gap of unknown length
* 99819 103856: contig of 4038 bp in length
* 103857 103956: gap of unknown length
* 103957 109114: contig of 5158 bp in length

```

Query Match 84.0% Score 16.8; DB 2; Length 119535;  
Best Local Similarity 90.0% Pred. No. 4.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTAAAGCCACGCTCTC 20  
Db 91492 ATATTAAAGCCACGCTCAC 91511

RESULT 7  
AC046134 168949 bp DNA linear PRI 15-MAY-2002  
LOCUS Homo sapiens 3 BAC RP11-553K23 (Rosewell Park Cancer Institute Human  
DEFINITION BAC library) complete sequence.  
AC046134  
AC046134.8 GI:20334471  
KEYWORDS HIG.

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 168949)  
Munzy D.M., Adams C., Adio-Oduola B., Ali-oshun, F.R., Allen C.,  
Alistbrooks S.L., Amaralunga H.C., Ayele M., Banks T.,  
Barbata J., Benton J., Bimege K., Blankenburg K., Bonin D.,  
Bouck J., Bowie S., Brieve M., Brown E., Brown M., Bryant N.P.,  
Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,  
Carroll T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,  
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,  
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,  
Devila M.L., Davis C., Davy-Carroll L., Dederich D.A.,  
Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,  
Douthwaite K.J., Draper H., Dugan-Pocha S., Durbin K.J.,  
Earnhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,  
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,  
Gabriel A., Geo J., Garcia A., Garner T., Garza N., Gill R.,  
Gottell J.H., Guevara M., Gunaratne P., Hale S., Hamilton K.,  
Harris C., Harris K., Hart M., Havlak P., Hawes A., He X.,  
Hernandez G., Hernandez O., Hodgson A., Hognes M., Holloway C.,  
Hollins B., Homsi F., Howard S., Huber J., Hulys S., Hume J.,  
Jackson L.E., Jacobson B., Jia Y., Johnson R., Jolyet S.,  
Jordah S., Karlsson E., Kelly S., Khan U., King L., Korah J.,  
Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,  
Lewis L., Li U., Li E., Lichtharge O., Lieu C., Liu J., Liu W.,  
Loulsegged H., Lozano R.J., Lu X., Lucier A., Lucier R., Luna R.,  
Ma J., Maheshwari M., Mapa P., Martin R., Martindale A.,  
Martinez E., Massey E., Mawhney E., McLeod M.P., Meador M.,  
Mei G., Metzker M., Miner G., Miner Z., Mitchell T., Mohabbat K.,  
Moore S., Morgan M., Moorish T., Morris S., Moser M., Neal D.,  
Nelson D., Newton J., Newton N., Nguyen A., Nguyen N., Nguyen N.,  
Nickerson E., Nwokenkwo S., Ogun M., Okwuon G., Oragunye N.,  
Oviedo R., Pace A., Payton B., Peery J., Peters L.,  
Pickers R., Primus E., Pu L.L., Quiles M., Ren Y., Rivers M.,  
Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savary G., Scherer S.,  
Scott G., Shen H., Shooshari N., Sisson I., Sodergren E.,  
Sonaike T., Sparks A., Stanley H., Stone H., Sutton A., Strack A.,  
Taber P., Tamerisa A., Tamerisa K., Tang H., Tansley J., Taylor C.,  
Taylor T., Telford B., Thomas N., Thomas S., Usmali K., Vasquez L.,  
Vera Y., Villalón D., Vinson R., Wang Q., Wang S., Ward-Moore S.,  
Warren R., Washington C., Watlington S., Williams G.,  
Williamson A., Wleczyk R., Wooden S., Worley K., Wu C., Wu Y.,  
Wu Y.F., Zhou J., Zorrilla S., Naylor S.L., Weinstein G. and  
Gibbs R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Direct Submission  
Unpublished  
2 (bases 1 to 168949)  
Worley K.C.  
Direct Submission  
Submitted (13-APR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 168949)  
Worley K.C.  
Direct Submission  
Submitted (26-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 168949)  
Worley K.C.  
Direct Submission  
Submitted (27-APR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 168949)  
Worley K.C.  
Direct Submission  
Submitted (15-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Apr 27, 2002 this sequence version replaced gi:19718590.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

Location/Qualifiers

1..168949

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="3"

/clone="RP11-553K23"

1..1999

/note="Overlaps bases 65964..67962 of clone AC110716"

/function="clone overlap"

/complement(388..505)

/rpt\_family="L2"

1269..1360

/rpt\_family="MIR"

1942..1985

/rpt\_family="(TG)n"

2142..2297

/rpt\_family="MIR"

3303..3347

/rpt\_family="CT-rich"

/complement(3360..3455)

/rpt\_family="L2"

/complement(6577..6808)

/rpt\_family="LTR16c"

/complement(7615..7924)

/rpt\_family="AluDb"

8780..9226

/rpt\_family="L2"

/complement(9397..9496)

/rpt\_family="MTR1G1"

9550..10657

/rpt\_family="L1MA4"

/complement(10685..11011)

/rpt\_family="MTR1G1"

11358..11391

/rpt\_family="(TC)n"

repeat\_region

repeat\_region 11542..11710

/rpt\_family="L2"

complement(42392..12606)

/rpt\_family="MTR1K"

complement(13485..13607)

/rpt\_family="MIR"

13574..13648

/rpt\_family="MIR"

13659..14141

/rpt\_family="MTR1D"

14143..14254

/rpt\_family="MIR"

complement(14657..14871)

/rpt\_family="MIR"

15168..15431

/rpt\_family="L1MC5"

complement(15432..15730)

/rpt\_family="AluDb"

15731..15760

/rpt\_family="L1MC5"

complement(17092..17266)

/rpt\_family="MER5A"

17276..17445

/rpt\_family="MER5B"

17452..17662

/rpt\_family="MER113"

17942..18329

/rpt\_family="L1ME2"

complement(20312..20585)

/rpt\_family="AluDb"

20615..20714

/rpt\_family="L2"

21914..22316

/rpt\_family="L2"

24589..24730

/rpt\_family="MIR"

26112..26136

/rpt\_family="AT-rich"

26537..26646

/rpt\_family="MIR"

27763..27794

/rpt\_family="(TTTG)n"

complement(28315..28354)

/rpt\_family="L1PA10"

complement(28725..29083)

/rpt\_family="TBE1C"

complement(29224..29691)

/rpt\_family="MTR1D"

Query Match 84.0%; Score 16.6; DB 9; Length 168949;

Best Local Similarity 90.0%; Pred. No. 4.2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AATTACGCCAGCCTCTC 20

Db 1233 ATATTACTCCACCTCTC 1252

#### RESULT 8

AC025890 184321 bp DNA linear HTG 24-AUG-2002

LOCUS Homo sapiens chromosome 3 clone RP11-553K23 map 3, WORKING DRAFT

DEFINITION SHDQNCB, 4 unordered pieces.

AC025890

AC025890.3 GI:8077134

VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (Bases 1 to 184321)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 3, clone RP11-553K23

JOURNAL  
REFERENCE  
AUTHORS

Unpublished  
2 (bases 1 to 184321)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,  
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,  
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kamp,L., Kartas,A.,  
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lebochzy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,  
McCarthy,M., McGowan,P., McGuirk,A., McKernan,K., McPheters,R.,  
Meltrin,J., Menus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Direct Submission  
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 184321)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,  
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,  
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kamp,L., Kartas,A.,  
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lebochzy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,  
McCarthy,M., McGowan,P., McGuirk,A., McKernan,K., McPheters,R.,  
Meltrin,J., Menus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7387411.  
All repeats were identified using RepeatMasker:  
Smith,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
-----  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
-----  
Project Information  
Center project name: L8244  
Center clone name: 553\_K\_23  
-----  
Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 181869 bases at least Q40  
Consensus quality: 183483 bases at least Q30  
Consensus quality: 183879 bases at least Q20  
Insert size: 21500; agarose-1p  
Insert size: 184021; sum-of-contents  
Quality coverage: 4.8 in Q20 bases; agarose-1p

Quality coverage: 5.6 in Q20 bases; sum-of-contents

NOTE: This is a 'working draft' sequence. It currently  
consists of 4 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 18607: contig of 18607 bp in length  
\* 18608 18707: gap of 100 bp  
\* 18708 53507: contig of 34800 bp in length  
\* 53508 53607: gap of 100 bp  
\* 53608 97651: contig of 44044 bp in length  
\* 97652 97751: gap of 100 bp  
\* 97752 184321: contig of 86570 bp in length.

FEATURES  
SOURCE

1. 184321

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="3"

/map="3"

/clone="RP11-553K23"

/clone.lib="RP11-11 Human Male BAC"

1. 18607

/note="assembly-fragment"

clone\_end:SP6

vector\_side:left"

18708..53507

/note="assembly-fragment"

53608..97651

/note="assembly-fragment"

97752..184321

/note="assembly-fragment"

clone\_end:T7

vector\_side:left"

51979 a 38476 c 39574 g 53992 t 300 others

BASP COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 2; Length 184321;

Best Local Similarity 90.0%; Pred. No. 4.2e+02;

Matches 18; Conservative 0; Mismatches 2; Gaps 0;

1 AATTAAGCCGACGCTC 20

Db 1239 AATTAAGCCGACGCTC 1258

RESULT 9

AK027375

LOCUS

DEFINITION

AK027375

ACCESSION

AK027375

KEYWORDS

SOURCE

ORGANISM

Hom sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Isogal,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,

Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,

Wagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,

Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,

Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,

Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,

Nakamura,Y., Nagahara,K., Masuno,Y., Niimiyu,K. and Iwayanagi,T.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 3226)



# TITLE JOURNAL COMMENT

George, R., Gonzalez, M., Guarin, H., Krommiller, B., Li, P., Liao, G., Miranda, A., Munhall, C. J., Munoo, J., Pacle, J., Paragas, V., Park, S., Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Rubin, G. M. and Celisner, S.  
Direct Submission  
Submitted (13-JUN-2002) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES source

gene  
CDS

BASE COUNT  
ORIGIN

1595 a 1434 c 1581 g 1268 t

Query Match 82.0%; Score 16.4; DB 3; Length 5878;  
Best Local Similarity 94.4%; Pred. No. 7.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AATACGCCACGCTCTC 20  
DB 2683 AATACGCCACGCTCTC 2666

## RESULT 12

AE009916/c 14086 bp DNA linear BCT 16-JAN-2002  
LOCUS Pyrobaculum aerophilum strain IM2 section 171 of 201 of the  
DEFINITION complete genome.  
ACCESSION AE009916 AE009441  
VERSION AE009916.1 GI:18161450  
KEYWORDS  
SOURCE  
ORGANISM

## REFERENCE

1 (bases 1 to 14086)  
Fitz-Gibbon, S.T., Ladner, H., Kim, U.-J., Stetter, K.O., Simon, M.I. and  
Miller, J.H.  
Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
aerophilum  
Proc. Natl. Acad. Sci. U.S.A. 99 (2), 984-989 (2002)

## AUTHORS

## JOURNAL

## FEATURES

## source

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

1.14086  
/organism="Pyrobaculum aerophilum"  
/strain="IM2"  
/db\_xref="taxon:13773"  
/gene="PAE3183"  
complement(118..231)  
/gene="PAE3183"  
/note="Hypothetical"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA164733.1"  
/db\_xref="GI:18161451"  
/translation="MPLSPALLSKRYKLVGTETELCGDYVSYCVYSG"  
270..322  
/gene="PAE3184"  
270..322  
/gene="PAE3184"  
/product="SRNA, predicted to direct ribose methylation of  
tRNA33 (1ys) at G10"  
366..1457  
/gene="PAE3184"  
366..1457  
/gene="PAE3184"  
/note="Protein fate: Protein modification and repair"  
/codon\_start=1  
/transl\_table=11  
/product="pyruvate formate-lyase activating enzyme  
homolog"  
/protein\_id="AA164734.1"  
/db\_xref="GI:18161452"  
/translation="MEAYISRAIPDGRVETCAACARCKLBEQGTGCGVRYRGGRIV  
LVYGLISALAIIDTEKKPLTHETPLFALYLSFTFGCWMACQICQNEELISQRRABGE  
EYTPELIVKAEISYGAGITVYNEPTLTFEFAADVGLAKSGLENTFVINGVTPPE  
AVKVASFELDAATVDFKGNADKEMFKKITYVDAEPIFETLAEKKRYGIWAEVVDLVV

```

gene
CDS
    EXVDDLEKARLRVRRVIDVDPVPIHLIRFHPDKMMDLPPPVETLEKHYEAAK
    EGARFAYGVNVPGRHYEYCPGCGRVVIRFRNGFDLLEINLYERKGEYRCKCGAKIP
    IKGIIMPTRWDEFRERYVYIQIFARVWR"
    /gene="PAE3185"
    2794..3171
    /note="Hypothetical; Conserved within genome"
    /codon_start=1
    /transl_table=11
    /product="PAREP9, associated with SRS8"
    /protein_id="AAL64735.1"
    /db_xref="GI:18161453"
    /translation="MTITGLVETRGVGYMILGKRLTYKSPRADTYTGLNANP
    LSIARLRLLPRKPDVRAAAYARALAAKVAEKCDSPWKVYTWELKAAVEDA
    ITYIOWIMPETRTIRFSRPPWMA"
    3475..4131
    /gene="PAE3187"
    3475..4131
    /note="Hypothetical; Conserved"
    /codon_start=1
    /transl_table=11
    /product="Conserved hypothetical protein"
    /protein_id="AAL64736.1"
    /db_xref="GI:18161454"
    /translation="MQRKRVSSGDLDELKERIDSHRLSDPSRLIQETIASRNLDK
    SAVDAVKITALERGEIVADYVDRAVAVPTVAVVYLIDNVDILMTIFA
    AKETIRGRHLTNDATYSIVSAELKEMLDIVPTVETIOTMTKKNADLRARRATIS
    SLERIDEIKENTLDKYALSLANVEPHYSLITYAKIKADNQDAHYALATVTVSY
    "
    4131..4895
    /gene="PAE3188"
    4131..4895
    /note="Hypothetical"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAL64737.1"
    /db_xref="GI:18161455"
    /translation="MELVGLIIMALLIYNNSLVLLGPTAMGAGIGAKARAVIAAAL
    LGLVLTWAQPISLPEEYIGALITLISLVKISLPIISVIGSIHARPPARVALMS
    PLVSAVYALCKTLKRRGPLALSLFLVMLTGFNNVALPEDALAAAVAGTGG
    LGSEKVIDIALARPKTAANVLYALGALAGILISVPISTLVAYGSIILAAVSQGM
    RIIRVEKPAKAVLATLALAAALIPYIKSLAPRA"
    complement(4852..5523)
    /gene="PAE3189"
    complement(4852..5523)
    /note="Hypothetical; Conserved"
    /codon_start=1
    /transl_table=11
    /product="Conserved hypothetical protein"
    /protein_id="AAL64738.1"
    /db_xref="GI:18161456"
    /translation="MPFSYGGELLISWGHDSFRITGGGVLYIDPYQLQVGDRAKDAI
    LITHEHFHCDPPSIQRIILKPSYVAVARAKAARNAVALISPGERREVGPKLI
    RAEPAYINKRDPARGVYVFRPKEDGRVAVLIEMGVRIIFRAGSDVFEFEREVRADY
    VLVPYGVYWTPOEAEEFVNAVMFKVALIPMRYGSIVASRRBAEEFKLVAPVQVVI
    LDREI"
    complement(5555..5728)
    /gene="PAE3192"
    complement(5555..5728)
    /note="Hypothetical; Conserved"
    /codon_start=1
    /transl_table=11
    /product="PAREP4"
    /protein_id="AAL64739.1"
    /db_xref="GI:18161457"
    /translation="MSKKQKLKFDIKAKQAFETDQYEVIEKQIARQPMFAYAKSPY
    TGIKIVRLGKKK"

gene
CDS
    5812..5901
    /gene="PAE3193"
    5812..5901
    /note="Hypothetical"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAL64740.1"
    /db_xref="GI:18161458"
    /translation="MDYGLALVLTALANSIHFPVRAGARR"
    5923..6342
    /gene="PAE3193a"
    5923..6342
    /note="Cellular processes; Adaptations to atypical
    conditions"
    /codon_start=1
    /transl_table=11
    /product="possible heat shock protein, hspx homolog part
    1, authentic frameshift"
    /protein_id="AAL64741.1"
    /db_xref="GI:18161459"
    /translation="MALIILAVANAVAVAPILFPVAVYIQAALGKIDYTAFAVLA
    AVSAAMVITISPYIINAVYGPDPDGLQQLVDVAAKLGGRVKAANVAVGPPKRLT
    RQLPVGVRGGDGLVKGHEPRRAGRRHRRAGTPIHK"
    6254..6838
    /gene="PAE3195"
    6254..6838
    /note="Cellular processes; Adaptations to atypical
    conditions"
    /codon_start=1
    /transl_table=11
    /product="possible heat shock protein, htpx homolog part
    2, authentic frameshift"
    /protein_id="AAL64742.1"
    /db_xref="GI:18161460"
    /translation="MAVTGILNTASRDELEAVYGHGELHINKNQMPITAGIISI
    AFTGVAIHLGLADREPSITAIAGVMAVVSFTVQLVLSFKLEETIRADGGR
    AAGKAMKAMRLAKIHKRYAQNPEALATSPKISGKALFLYALINAAANPLDITPEV
    RQMLKIQSWIDELTSHPPIPKRLVLSAAP"
    6928..7006
    /gene="PAE3194"
    6928..7006
    /note="tRNA-gly"
    /product="tRNA-gly"
    7139..8140
    /gene="PAE3196"
    7139..8140
    /note="Hypothetical"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAL64743.1"
    /db_xref="GI:18161461"
    /translation="MVEICRSLLCGLGAGRSQRLYYAAGLVDEGLIAASRADEL
    RVLVDSGVPRPRVRYRMEVRYRVPSSHAKFTIPEDAGPSAALVGSADLTAGLNG
    NIEAVVILRGGAARPLADMEFNLIRALPLTEQYVADWEGPEEALRKMGVAVKANE
    RLAEITLGVSAHCSRBDPLNCARLYARAVRRECGGDPENCARATGYSARALLSA
    PSAYILAGHYGCARALAAARLIEGVGRIDSGMEVAAVVOAGASCGEAKRAAEE
    IRRUDSYRDNTRYRNPITPYRLIFLAKMILPRTGCRILIGEVYTKRQVARENELIC"
    complement(8131..10143)
    /gene="PAE3197"
    complement(8131..10143)

Query Match      82.0%; Score 16.4; DB 1; Length 14086;
Best Local Similarity 94.4%; Pred. No. 7.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 ATATTACGCCCGCCTC 18
|||||

```

Db 5813 ANATTACGCCACGCC 5796

RESULT 13  
AC014130/c  
LOCUS  
DEFINITION AC014130 40123 bp DNA linear HTG 16-NOV-1999  
Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
pieces.  
ACCESSION AC014130 GI:6437205  
VERSION  
KEYWORDS HTG, HTGS\_PHASE2.  
SOURCE  
ORGANISM Drosophila melanogaster.  
Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 40123)  
REFERENCE  
AUTHORS Adams, M. and Venter, J.C.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT This sequence was identified as CDM:10210776 by the submitter.  
For further information on this sequence e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source 1.40123  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 10427 a 9574 c 9411 g 10711 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 2; Length 40123;  
Best Local Similarity 94.4%; Pred. No. 7.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTACGCCACGCTCTC 20  
|||||||  
Db 39472 ATTACGCCCGCTCTC 39455

RESULT 14  
AC016523 65961 bp DNA linear HTG 02-DEC-1999  
LOCUS  
DEFINITION AC016523 65961 bp DNA linear HTG 02-DEC-1999  
Drosophila melanogaster chromosome X clone BACR40P05 (D1208)  
RPCI-98 40.P.5 map 16B-16B strain Y; cn bw sp, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 60 unordered pieces.  
ACCESSION AC016523  
VERSION  
KEYWORDS HTG, HTGS\_PHASE1.  
SOURCE Drosophila melanogaster.  
Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 65961)  
REFERENCE  
AUTHORS Celniker, S.E., Acharyani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,  
Butenoff, C., Chapoy, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,  
Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,  
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,  
Richards, S., Sethi, H., Switskas, R.R., Wan, K.H., Webster, D.,  
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.  
TITLE Sequencing of Drosophila melanogaster  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 65961)  
AUTHORS Celniker, S.E., Acharyani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,  
Butenoff, C., Chapoy, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,  
Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,  
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,  
Richards, S., Sethi, H., Switskas, R.R., Wan, K.H., Webster, D.,  
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.

TITLE JOURNAL  
COMMENT

Submitted (02-DEC-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 60 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
657 736: contig of 656 bp in length  
737 736: gap of unknown length  
1733 1733: contig of 997 bp in length  
1734 1813: gap of unknown length  
1813 2418: contig of 603 bp in length  
2418 2498: gap of unknown length  
2498 3364: contig of 866 bp in length  
3364 3444: gap of unknown length  
3444 4009: contig of 565 bp in length  
4009 4089: gap of unknown length  
4089 4784: contig of 695 bp in length  
4784 4864: gap of unknown length  
4864 4785: gap of unknown length  
4785 5717: contig of 853 bp in length  
5717 5797: gap of unknown length  
5797 6515: contig of 718 bp in length  
6515 6595: gap of unknown length  
6595 7319: contig of 724 bp in length  
7319 7320: gap of unknown length  
7320 7400: gap of unknown length  
7400 8205: contig of 806 bp in length  
8205 8285: gap of unknown length  
8285 8954: contig of 669 bp in length  
8954 9034: gap of unknown length  
9034 9035: gap of unknown length  
9035 10372: contig of 1337 bp in length  
10372 10451: gap of unknown length  
10451 11326: contig of 785 bp in length  
11326 11316: gap of unknown length  
11316 12306: contig of 990 bp in length  
12306 12386: gap of unknown length  
12386 13131: contig of 745 bp in length  
13131 13332: gap of unknown length  
13332 13387: gap of unknown length  
13387 14351: contig of 1140 bp in length  
14351 14352: gap of unknown length  
14352 15050: contig of 619 bp in length  
15050 15130: gap of unknown length  
15130 16538: contig of 1328 bp in length  
16538 16539: gap of unknown length  
16539 17995: contig of 1457 bp in length  
17995 18075: gap of unknown length  
18075 18076: gap of unknown length  
18076 19235: contig of 1160 bp in length  
19235 19315: gap of unknown length  
19315 21063: contig of 1748 bp in length  
21063 21143: gap of unknown length  
21143 22659: contig of 1516 bp in length  
22659 22739: gap of unknown length  
22739 24028: contig of 1268 bp in length  
24028 24108: gap of unknown length  
24108 24779: contig of 671 bp in length  
24779 24859: gap of unknown length  
24859 26341: contig of 1482 bp in length  
26341 26421: gap of unknown length  
26421 27687: contig of 1266 bp in length  
27687 27688: gap of unknown length  
27688 29294: contig of 1527 bp in length

```

* 29295 29374: gap of unknown length
* 29375 30634: contig of 1260 bp in length
* 30635 30714: gap of unknown length
* 30715 32904: contig of 2190 bp in length
* 32905 32984: gap of unknown length
* 32985 34623: contig of 1639 bp in length
* 34624 34703: gap of unknown length
* 34704 36780: contig of 2077 bp in length
* 36781 36860: gap of unknown length
* 36861 38926: contig of 2066 bp in length
* 38927 39006: gap of unknown length
* 39007 41830: contig of 2824 bp in length
* 41831 41910: gap of unknown length
* 41911 45034: contig of 3124 bp in length
* 45035 45114: gap of unknown length
* 45115 49237: contig of 4123 bp in length
* 49238 49317: gap of unknown length
* 49318 49969: contig of 652 bp in length
* 49970 50049: gap of unknown length
* 50050 50591: contig of 642 bp in length
* 50592 50771: gap of unknown length
* 50772 51393: contig of 622 bp in length
* 51394 51473: gap of unknown length
* 51474 52111: contig of 638 bp in length
* 52112 52836: contig of 645 bp in length
* 52837 52916: gap of unknown length
* 52917 53464: contig of 548 bp in length
* 53465 53544: gap of unknown length
* 53545 53758: contig of 214 bp in length
* 53759 53838: gap of unknown length
* 53839 54430: contig of 592 bp in length
* 54431 54510: gap of unknown length
* 54511 55156: contig of 646 bp in length
* 55157 55236: gap of unknown length
* 55237 55863: contig of 627 bp in length
* 55864 55943: gap of unknown length
* 55944 56516: contig of 573 bp in length
* 56517 56596: gap of unknown length
* 56597 57180: contig of 584 bp in length
* 57181 57260: gap of unknown length
* 57261 57872: contig of 612 bp in length
* 57873 57952: gap of unknown length
* 57953 58653: contig of 701 bp in length
* 58654 58733: gap of unknown length
* 58734 59364: contig of 631 bp in length
* 59365 59444: gap of unknown length
* 59445 60126: contig of 682 bp in length
* 60127 60206: gap of unknown length
* 60207 60835: contig of 629 bp in length
* 60836 60915: gap of unknown length
* 60916 61535: contig of 620 bp in length
* 61536 61615: gap of unknown length
* 61616 62021: contig of 406 bp in length
* 62022 62101: gap of unknown length
* 62102 62937: contig of 836 bp in length
* 62938 63017: gap of unknown length
* 63018 63659: contig of 642 bp in length
* 63660 63739: gap of unknown length
* 63740 64372: contig of 633 bp in length
* 64373 64452: gap of unknown length
* 64453 65031: contig of 579 bp in length
* 65032 65111: gap of unknown length
* 65112 65473: contig of 362 bp in length
* 65474 65563: gap of unknown length
* 65564 65961: contig of 408 bp in length.

```

## FEATURES

source

```

1. 65961
   /organism="Drosophila melanogaster"
   /strain="y; cn bw sp"
   /db_xref="taxon:7227"
   /chromosome="X"
   /map="16B-16B"

```

```

BASE COUNT      16392 a 14437 c 13716 g 16689 t 4727 others
ORIGIN

```

```

Query Match      82.0%; Score 16.4; DB 2; Length 65961;
Best Local Similarity 94.4%; Pred. No. 7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

GY      3 ATTAACGCCACGCTCTC 20
DB      50448 ATTAACGCCACGCTCTC 50465

```

```

RESULT 15
AC012649/c      80866 bp   DNA   linear   HTG 04-NOV-1999
LOCUS      Drosophila melanogaster chromosome X clone BACR07N08 (D1121)
DEFINITION  RPCI-98 07.N.8 map 17D-17E strain y; cn bw sp, *** SEQUENCING IN
PROGRESS ***. 101 unordered pieces.

```

```

ACCESSION  AC012649.2 GI:6226990
VERSION    AC012649
KEYWORDS   HTG; HTGS; PHASE1.
SOURCE     Drosophila melanogaster.
ORGANISM   Drosophila melanogaster.

```

```

REFERENCE   1 (bases 1 to 80866)
AUTHORS    Butenheft,C., Agbayani,A., Arcalana,T.T., Baxter,E., Blazey,R.G.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,K.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Kara,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.

```

```

TITLE      Sequencing of Drosophila melanogaster
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 80866)
AUTHORS    Celniker,S.E., Agbayani,A., Arcalana,T.T., Baxter,E., Blazey,R.G.,
            Butenheft,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Butenheft,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.R., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Kara,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomontan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,U., Sequeira,A., Sethi,H., Snit,E.,
            Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.

```

## TITLE

```

JOURNAL    Submitted (02-NOV-1999) Drosophila genome center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA

```

## COMMENT

```

On Nov 4, 1999 this sequence version replaced g1:6175114.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 101 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1      577: contig of 577 bp in length
*      578      657: gap of unknown length
*      658      1404: contig of 747 bp in length
*      1405      1484: gap of unknown length
*      1485      2313: contig of 829 bp in length

```

*	2314	2393: gap of unknown length
*	2394	3131: contig of 738 bp in length
*	3132	3211: gap of unknown length
*	3212	3719: contig of 508 bp in length
*	3720	3799: gap of unknown length
*	3800	4897: contig of 1098 bp in length
*	4898	4977: gap of unknown length
*	4978	5799: contig of 822 bp in length
*	5800	5879: gap of unknown length
*	5880	6469: contig of 590 bp in length
*	6470	6549: gap of unknown length
*	6550	7442: contig of 893 bp in length
*	7443	7522: gap of unknown length
*	7523	8486: contig of 964 bp in length
*	8487	8566: gap of unknown length
*	8567	9246: contig of 680 bp in length
*	9247	9326: gap of unknown length
*	9327	9971: contig of 645 bp in length
*	9972	10051: gap of unknown length
*	10052	10632: contig of 581 bp in length
*	10633	10712: gap of unknown length
*	10713	11328: contig of 616 bp in length
*	11329	11408: gap of unknown length
*	11409	12070: contig of 662 bp in length
*	12071	12150: gap of unknown length
*	12151	12954: contig of 804 bp in length
*	12955	13034: gap of unknown length
*	13035	13887: contig of 853 bp in length
*	13888	13967: gap of unknown length
*	13968	14675: contig of 708 bp in length
*	14676	14755: gap of unknown length
*	14756	15777: contig of 1022 bp in length
*	15778	15857: gap of unknown length
*	15858	16561: contig of 704 bp in length
*	16562	16641: gap of unknown length
*	16642	17656: contig of 1015 bp in length
*	17657	17736: gap of unknown length
*	17737	18347: contig of 611 bp in length
*	18348	18427: gap of unknown length
*	18428	19269: contig of 842 bp in length
*	19270	19349: gap of unknown length
*	19350	20123: contig of 774 bp in length
*	20124	20203: gap of unknown length
*	20204	21037: contig of 834 bp in length
*	21038	21117: gap of unknown length
*	21118	21880: contig of 763 bp in length
*	21881	21960: gap of unknown length
*	21961	23063: contig of 1103 bp in length
*	23064	23143: gap of unknown length
*	23144	23742: contig of 599 bp in length
*	23743	23822: gap of unknown length
*	23823	24697: contig of 875 bp in length
*	24698	24777: gap of unknown length
*	24778	25699: contig of 922 bp in length
*	25700	25779: gap of unknown length
*	25780	26547: contig of 768 bp in length
*	26548	26627: gap of unknown length
*	26628	27682: contig of 1055 bp in length
*	27683	27762: gap of unknown length
*	27763	28171: contig of 409 bp in length
*	28172	28251: gap of unknown length
*	28252	28823: contig of 572 bp in length
*	28824	28903: gap of unknown length
*	28904	29965: contig of 1062 bp in length
*	29966	30045: gap of unknown length
*	30046	31173: contig of 1128 bp in length
*	31174	31253: gap of unknown length
*	31254	32156: contig of 903 bp in length
*	32155	32236: gap of unknown length
*	32237	32872: contig of 636 bp in length
*	32873	32952: gap of unknown length
*	32953	33817: contig of 865 bp in length
*	33818	33897: gap of unknown length

---

*	33898	35066: contig of 1169 bp in length
*	35067	35146: gap of unknown length
*	35147	37188: contig of 2042 bp in length
*	37189	37268: gap of unknown length
*	37269	38406: contig of 1138 bp in length
*	38407	38486: gap of unknown length
*	38487	40075: contig of 1589 bp in length
*	40076	40155: gap of unknown length
*	40156	41204: contig of 1049 bp in length
*	41205	41284: gap of unknown length
*	41285	42613: contig of 1329 bp in length
*	42614	42693: gap of unknown length
*	42694	45350: contig of 2657 bp in length
*	45351	45430: gap of unknown length
*	45431	45914: contig of 484 bp in length
*	45915	45994: gap of unknown length
*	45995	46589: contig of 595 bp in length
*	46590	46669: gap of unknown length
*	46670	47222: contig of 553 bp in length
*	47223	47302: gap of unknown length
*	47303	47887: contig of 585 bp in length
*	47888	47967: gap of unknown length
*	47968	48553: contig of 586 bp in length
*	48554	48633: gap of unknown length
*	48634	49266: contig of 633 bp in length
*	49267	49346: gap of unknown length
*	49347	49944: contig of 598 bp in length
*	49945	50024: gap of unknown length
*	50025	50642: contig of 618 bp in length
*	50643	50722: gap of unknown length
*	50723	51227: contig of 505 bp in length
*	51228	51307: gap of unknown length
*	51308	52034: contig of 727 bp in length
*	52035	52114: gap of unknown length
*	52115	52686: contig of 572 bp in length
*	52687	52766: gap of unknown length
*	52767	53363: contig of 597 bp in length
*	53364	53443: gap of unknown length
*	53444	53993: contig of 550 bp in length
*	53994	54073: gap of unknown length
*	54074	54434: contig of 361 bp in length
*	54435	54514: gap of unknown length
*	54515	55129: contig of 615 bp in length
*	55130	55209: gap of unknown length
*	55210	55887: contig of 678 bp in length
*	55888	55967: gap of unknown length
*	55968	56510: contig of 543 bp in length
*	56511	56590: gap of unknown length
*	56591	57022: contig of 432 bp in length
*	57023	57102: gap of unknown length
*	57103	57394: contig of 292 bp in length
*	57395	57474: gap of unknown length
*	57475	58106: contig of 632 bp in length
*	58107	58186: gap of unknown length
*	58187	58920: contig of 734 bp in length
*	58921	59000: gap of unknown length
*	59001	59609: contig of 609 bp in length
*	59610	59689: gap of unknown length
*	59689	60571: contig of 682 bp in length
*	60372	60451: gap of unknown length
*	60452	61001: contig of 550 bp in length
*	61002	61081: gap of unknown length
*	61082	61556: contig of 575 bp in length
*	61557	61736: gap of unknown length
*	61737	62261: contig of 525 bp in length
*	62262	62341: gap of unknown length
*	62342	62865: contig of 524 bp in length
*	62866	62945: gap of unknown length
*	62946	63538: contig of 593 bp in length
*	63539	63618: gap of unknown length
*	63619	64136: contig of 518 bp in length
*	64137	64216: gap of unknown length

Query Match 82.0%; Score 16.4; DB 2; Length 80866;  
Best Local Similarity 94.4%; Pred. No. 7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTAAAGCCGACGCTC 20  
|||||  
DB 62105 ATTAAAGCCGACGCTC 62088

Search completed: May 12, 2003, 01:53:59  
Job time : 244 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:45:59 ; Search time 35.2143 Seconds  
(without alignments)  
1279.025 Million cell updates/sec

Title: US-09-980-277-3  
Perfect score: 20  
Sequence: 1 atataacgcccacgcctc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match length	ID	Description
1	20	100.0	20	AAIT0402
2	20	100.0	120	AAIT0400
3	20	100.0	21	AAA46170
4	16.8	84.0	747	ABO22068
5	16.8	84.0	747	ABO22069
6	16.8	84.0	748	ABO53094
7	16.8	84.0	748	ABO53095
8	16.4	82.0	3226	AAH18430
9	16.4	82.0	3251	ABL13875

10	16.4	82.0	5484	ABL13874
11	15.8	79.0	729	AA566682
12	15.8	79.0	729	AA575819
13	15.8	79.0	890	ABO34482
14	15.8	79.0	890	ABO34483
15	15.8	79.0	3759	AB105751
16	15.8	79.0	6020	AB105750
17	15.8	79.0	6075	AB133551
18	15.8	79.0	6207	AA563341
19	15.8	79.0	6922	AA563533
20	15.8	79.0	6922	AB170379
21	15.8	79.0	6922	AA561331
22	15.8	79.0	6922	ABK31410
23	15.8	79.0	7631	ABK32860
24	15.8	79.0	8076	ABK39954
25	15.8	79.0	9543	ABNR0049
26	15.8	79.0	19459	ABK10528
27	15.8	79.0	19459	ABK31213
28	15.8	79.0	33053	ABO67006
29	15.8	79.0	34548	AB170604
30	15.4	77.0	336	AA567489
31	15.4	77.0	2535	AA553763
32	15.4	77.0	2547	AA567908
33	15.4	77.0	2577	AA565651
34	15.4	77.0	2577	AA533652
35	15.4	77.0	2627	AA514411
36	15.4	77.0	2627	AA552087
37	15.2	76.0	103	ABV14698
38	15.2	76.0	105	ABV03444
39	15.2	76.0	124	AA123967
40	15.2	76.0	130	AA115670
41	15.2	76.0	139	ABV03603
42	15.2	76.0	144	ABV05529
43	15.2	76.0	149	ABV07191
44	15.2	76.0	165	ABV05266
45	15.2	76.0	179	AA119730

## ALIGNMENTS

RESULT 1  
AAIT0402  
ID AAIT0402 standard; DNA; 20 BP.  
XX  
AC AAIT0402:  
XX  
DT 07-JAN-2002 (first entry)  
XX  
DE Schistosoma mansoni repeater unit 3' PCR primer.  
XX  
KW Schistosomiasis; infection; diagnosis; PCR primer; ss.  
XX  
OS Schistosoma mansoni.  
XX  
PN WO200175148-A1.  
XX  
PD 11-OCT-2001.  
XX  
PF 04-APR-2001; 2001WO-BR00035.  
XX  
PR 04-APR-2000; 2000BR-0001536.  
XX  
PA (FIOC-) FIOCRUZ FUNDACAO CRUZ OSWALDO.  
XX  
PI Teles Rabello AL, Dias Neto E, Pontes JA;  
XX  
DR WPI; 2001-648561/74.  
XX  
PT Diagnosing infection by parasites of the Schistosoma sp. in a  
PT biological sample especially useful in cases of low infection  
PT intensity; comprises detecting a specific region of the DNA of  
PT Schistosoma by polymerase chain reaction -

Drosophila melanog  
DNA encoding novel  
DNA encoding novel  
Oligonucleotide to  
Oligonucleotide to  
Drosophila melanog  
Drosophila melanog  
Human immune syste  
Chemically pretrea  
Tumour suppressor  
Chemically treated  
Human gene regulat  
Signal transductio  
Human immune syste  
Human chemically m  
Chemically treated  
Signal transductio  
Human angiotensin  
Chemically treated  
H. pylori cytoplas  
Helicobacter pylor  
H. pylori HP048 e  
Helicobacter pylor  
H. pylori GHP0 615  
Human prostate exp  
Human prostate exp  
Human breast cance  
Human breast cance  
Human prostate exp  
Human prostate exp  
Human prostate exp  
Human prostate exp  
Human breast cance

XX Claim 4; Page 13; 34pp; English.  
PS The present sequence is that of a claimed oligonucleotide primer.  
XX derived from the 3' region of a highly repeated short DNA sequence  
CC (see AAT70400) of the Schistosoma mansoni genome. This 3' primer,  
CC and the 5' primer given in AAT70401, are used in the method of the  
CC invention for diagnosing Schistosoma infection. The method involves  
CC collection of the sample to be examined, extraction of Schistosoma  
CC sp. DNA, amplification by PCR, separation of PCR products by  
CC electrophoresis, and detection e.g. by colouring with silver salts.  
CC A kit for diagnosing infection, which includes the primers, is  
CC claimed. The method is useful for detecting Schistosoma sp.  
CC parasites by detecting the parasite's DNA in a biological sample.  
CC It is especially useful in cases of low infection intensity for  
CC which parasitological stool tests demonstrate little sensitivity.  
SQ Sequence 20 BP; 5 A; 8 C; 2 G; 5 T; 0 other;  
Query Match 100.0%; Score 20; DB 22; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.38;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AATTTACGCCACGCTCTC 20  
DB 1 AATTTACGCCACGCTCTC 20  
|||||  
RESULT 2  
AAT70400/C  
ID AAT70400 standard; DNA; 120 BP.  
XX  
AC AAT70400;  
XX  
DT 07-JAN-2002 (first entry)  
XX  
DE Schistosoma mansoni repeater unit DNA.  
XX  
KW Schistosomiasis; infection; diagnosis; ds.  
XX  
OS Schistosoma mansoni.  
XX  
FH Key Location/Qualifiers  
FT primer\_bind /tag= a  
FT complement (1..19)  
FT /note= "primer of AAT70401"  
FT primer\_bind 91..110  
FT /tag= b  
FT /note= "primer of AAT70402"  
XX  
PN WO200175148-A1.  
XX  
PD 11-OCT-2001.  
XX  
PF 04-APR-2001; 2001MO-BR00035.  
XX  
PR 04-APR-2000; 2000BR-0001536.  
XX  
PA (FIOC-) FIOCRUZ FUNDACAO CRUZ OSWALDO.  
XX  
PI Teles Rabello AL, Dias Neto E, Pontes LA;  
XX  
DR WPI; 2001-648561/74.  
XX  
PT Diagnosing infection by parasites of the Schistosoma sp. in a  
CC biological sample especially useful in cases of low infection  
PT intensity, comprises detecting a specific region of the DNA of  
PT Schistosoma by polymerase chain reaction -  
XX  
PS Claim 1; Page 22; 34pp; English.  
XX  
CC The present sequence is that of a highly repeated short DNA  
CC sequence in the genome of Schistosoma mansoni. PCR primers

CC (see AAT70401 and AAT70402) that flank this sequence are used in  
CC the method of the invention for diagnosing Schistosoma infection.  
CC The method involves collection of the sample to be examined,  
CC extraction of Schistosoma sp. DNA, amplification by PCR, separation  
CC of PCR products by electrophoresis, and detection e.g. by colouring  
CC with silver salts. A kit for diagnosing infection, which includes  
CC the primers, is claimed. The method is useful for detecting  
CC Schistosoma sp. parasites by detecting the DNA of the parasite in a  
CC biological sample. It is especially useful in cases of low  
CC infection intensity for which parasitological stool tests  
CC demonstrate little sensitivity.  
SQ Sequence 120 BP; 36 A; 20 C; 26 G; 38 T; 0 other;  
Query Match 100.0%; Score 20; DB 22; Length 120;  
Best Local Similarity 100.0%; Pred. No. 0.46;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AATTTACGCCACGCTCTC 20  
DB 110 AATTTACGCCACGCTCTC 91  
|||||  
RESULT 3  
AAA6170/C  
ID AAA6170 standard; DNA; 6494 BP.  
XX  
AC AAA6170;  
XX  
DT 27-SEP-2000 (first entry)  
XX  
DE GFP-Sm1-7 fusion protein construct.  
XX  
KW GFP-Sm1-7 fusion construct; circular; green fluorescent protein;  
KW glutathione S-transferase; eukaryotic diploid multicellular parasite;  
KW universal graft; transgenic eukaryotic parasite; acquired deficiency;  
KW genetic deficiency; hormone deficiency; metabolic deficiency;  
KW haematological deficiency; immunological deficiency; immunotherapy;  
KW anti-microbial therapy; anti-cancer therapy; drug addiction;  
KW poisoning condition; geriatric condition; ds.  
XX  
OS Synthetic.  
XX  
PN WO200032804-A1.  
XX  
PD 08-JUN-2000.  
XX  
PF 01-DEC-1999; 99WO-IL00651.  
XX  
PR 01-DEC-1998; 98US-0201850.  
XX  
PA (YISS ) YISSUM RRS & DEV CO.  
XX  
PI Hamburger J, Laban A;  
XX  
DR WPI; 2000-412348/35.  
XX  
PT Eukaryotic diploid multicellular parasite useful as universal grafts  
PT for in vivo delivery of beneficial gene products in humans and animals  
PT involves transformation with a transgene -  
XX  
PS Example 2; Fig 6; 90pp; English.  
XX  
CC This sequence represents a GFP-Sm1-7 fusion construct contained within a  
CC recombinant vector. This sequence contains the green fluorescent protein  
CC (GFP) coding sequence from *Aequorea victoria*, and the  
CC Sm1-7 coding sequence from *Schistosoma*, along with  
CC promoter sequences. The invention relates to a eukaryotic diploid  
CC multicellular parasite transformed with a transgene. Transgenic  
CC eukaryotic parasites are useful as universal grafts for in vivo delivery  
CC of beneficial gene product in humans and animals. The parasites can  
CC particularly be used for restoration of deficiencies whether acquired or  
CC genetic, such as hormone deficiencies, metabolic deficiencies,

CC haematological deficiencies, immunological deficiencies, immunotherapy,  
 CC anti-microbial therapy, anti-cancer therapy. They can also be used for  
 CC treatment of drug addiction, of poisoning conditions, and for  
 CC amelioration of geriatric conditions. Treatment of humans with in vivo  
 CC transgenes are universally compatible, readily available and inexpensive.  
 CC genotypic alterations of the patients is avoided, reducing risks of  
 CC mutagenesis and malignant transformation.

SQ Sequence 6494 BP; 1754 A; 1373 C; 1481 G; 1886 T; 0 other;

Query Match 100.0%; Score 20; DB 21; Length 6494;

Best Local Similarity 100.0%; Pred. No. 0.7; Mismatches 0; Gaps 0;

Matches 20; Conservative 0; Indels 0; Gaps 0;

QY 1 ATATTAAAGCCACGCTCTC 20  
 |||||  
 Db 110 ATATTAAAGCCACGCTCTC 91

#### RESULT 4

ABQ22068/C  
 ID ABQ22068 standard; DNA; 747 BP.

AC ABQ22068;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 8659.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIDENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Gueitig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the

CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

SQ Sequence 747 BP; 99 A; 102 C; 297 G; 249 T; 0 other;

Query Match 84.0%; Score 16.8; DB 24; Length 747;

Best Local Similarity 90.0%; Pred. No. 28; Mismatches 2; Indels 0; Gaps 0;

Matches 18; Conservative 0; Indels 0; Gaps 0;

QY 1 ATATTAAAGCCACGCTCTC 20  
 |||||  
 Db 400 ATATTAAAGCCACGCTCTC 381

#### RESULT 5

ABQ22069  
 ID ABQ22069 standard; DNA; 747 BP.

AC ABQ22069;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 8660.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIDENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Gueitig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in

SQ Sequence 748 BP; 104 A; 101 C; 282 G; 261 T; 0 other;

Query Match	84.0%;	Score 16.8;	DB 24;	Length 748;
Best Local Similarity	90.0%;	Pred. No. 28;		

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AATTAAAGCCCAAGCTCTC 20  
 |||||  
 Db 439 AATTAAAGCCCAAGCTCTC 458

RESULT 8  
 ABL18430/c  
 ID ABL18430 standard; cDNA; 3226 BP.

AC ABL18430;

DF 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:18513.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EPI074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

XX Claim 8; SEQ ID 18513; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAB03166 to AAB13628 and

CC AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAB13629 to AAB13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

Best Local Similarity 94.4%; Pred. No. 53;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AATTAAAGCCCAAGCTC 18  
 |||||  
 Db 615 AATTAAAGCCCAAGCTC 598

RESULT 9  
 ABL13875/c  
 ID ABL13875 standard; cDNA; 3251 BP.

AC ABL13875;

DF 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36107.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB69772.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 36107; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-AB130511), expressed DNA

CC sequences (ABL1840-ABL16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3251 BP; 849 A; 820 C; 880 G; 702 T; 0 other;

XX Query Match 82.0%; Score 16.4; DB 23; Length 3251;

XX Best Local Similarity 94.4%; Pred. No. 53;

XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

XX QY 3 AATTAAAGCCCAAGCTCTC 20

XX |||||

XX Db 2810 AATTAAAGCCCAAGCTCTC 2793

XX RESULT 10

XX ABL13874

XX ID ABL13874 standard; cDNA; 5484 BP.

XX AC ABL13874;

XX 26-MAR-2002 (first entry)  
DN  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 36104.  
DE  
XX  
XX Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
PN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
PR  
XX  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PMD, Myers EW;  
PI  
XX  
XX WPI; 2001-656860/75.  
DR  
XX  
XX P-PSDB; ABB69771.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
PT  
XX  
XX Claim 1; SEQ ID NO 36104; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-AB101875), expressed DNA  
CC sequences (AB101840-AB101875) and the encoded proteins  
CC (AB101840-AB101875).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
XX Sequence 5484 BP; 1402 A; 1280 C; 1322 G; 1480 T; 0 other;  
SQ  
XX  
XX Query Match 82.0%; Score 16.4; DB 23; Length 5484;  
Best Local Similarity 94.4%; Pred. No. 56;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 3 ATTAAGCCGACGCTCTC 20  
DB 652 ATTAAGCCGACGCTCTC 669  
|||||  
RESULT 11  
AAS66682/c  
ID AAS66682 standard; cDNA; 729 BP.  
XX  
XX AAS66682;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DN  
XX  
XX DNA encoding novel human diagnostic protein #2486.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD

XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR  
XX  
XX 23-AUG-2000; 2000US-0649167.  
PR  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR  
XX  
XX P-PSDB; ABB02495.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PT  
XX  
XX Claim 1; SEQ ID NO 2486; 103pp; English.  
PS  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
XX Sequence 729 BP; 269 A; 175 C; 148 G; 137 T; 0 other;  
SQ  
XX  
XX Query Match 79.0%; Score 15.8; DB 23; Length 729;  
Best Local Similarity 89.5%; Pred. No. 94;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 ATATTACGCCGACGCTCT 19  
DB 514 ATATTACGCCGACGCTCT 496  
|||||  
RESULT 12  
AAS75819/c  
ID AAS75819 standard; cDNA; 729 BP.  
XX  
XX AAS75819;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DN  
XX  
XX DNA encoding novel human diagnostic protein #11623.  
DE  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
PF

XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
DR P-PSDB; ABG11632.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID NO 11623; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 729 BP; 269 A; 175 C; 148 G; 137 T; 0 other;  
Query Match 79.0%; Score 15.8; DB 23; Length 729;  
Best Local Similarity 89.5%; Pred. No. 94;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATATTACGCCCGCTCT 19  
DB 514 ATATTACGCCCGCTCT 496  
RESULT 13  
ABQ34482/C  
ID ABQ34482 standard; DNA; 890 BP.  
XX  
AC ABQ34482;  
XX  
DE 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21073.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI: 2002-371829/40.  
DR  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ34410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.  
XX  
SQ Sequence 890 BP; 134 A; 106 C; 283 G; 367 T; 0 other;  
Query Match 79.0%; Score 15.8; DB 24; Length 890;  
Best Local Similarity 89.5%; Pred. No. 96;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 TATTACGCCCGCTCTC 20  
DB 23 TATTACGCCCGCTCTC 5  
RESULT 14  
ABQ34483  
ID ABQ34483 standard; DNA; 890 BP.  
XX  
AC ABQ34483;  
XX  
DE 12-JUL-2002 (first entry)  
XX  
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 21074.  
XX  
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
KW drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200218632-A2.  
XX  
PD 07-MAR-2002.  
XX  
PF 01-SEP-2001; 2001WO-EP10074.  
XX  
PR 01-SEP-2000; 2000DE-1043826.  
PR 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.  
 PA  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 DR WPI: 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12: 56pp + Sequence Listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C<sub>5</sub> to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SQ Sequence 890 BP; 367 A; 283 C; 106 G; 134 T; 0 other;  
 Query Match 79.0%; Score 15.8; DB 24; Length 890;  
 Best Local Similarity 89.5%; Pred. No. 96;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 2 TATTAAAGCCGACGCTC 20  
 DB 868 TCTTAAGCCGACGCTC 886  
 RESULT 15  
 ABL05751  
 ID ABL05751 standard; cDNA; 3759 BP.  
 XX  
 AC ABL05751;  
 XX  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11735.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN MO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NT.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX

DR WPI: 2001-656860/75.  
 DR P-RSDB; ABB61648.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1: SEQ ID NO 11735; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABP30511), expressed DNA  
 CC sequences (ABP57737-ABP72072).  
 CC (ABP57737-ABP72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pt\_sequences.  
 XX  
 SQ Sequence 3759 BP; 963 A; 927 C; 946 G; 923 T; 0 other;  
 Query Match 79.0%; Score 15.8; DB 23; Length 3759;  
 Best Local Similarity 89.5%; Pred. No. 11e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 OY 2 TATTAAAGCCGACGCTC 20  
 DB 3044 TATTAAAGCCGACGCTC 3062

Search completed: May 12, 2003, 01:14:58  
 Job time : 37.2143 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:08:23 ; Search time 8.78571 Seconds  
(without alignments)  
698.126 Million cell updates/sec

Title: US-09-980-277-3

Perfect score: 20

Sequence: 1 atattacgcccaagctctc 20

Scoring table: IDENTITY\_MTC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents.NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*

2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*

3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*

4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*

5: /cgn2\_6/ptodata/1/ina/PCFUS.COMB.seq.\*

6: /cgn2\_6/ptodata/1/ina/Backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	14.8	74.0	3636	3 US-09-074-579-2	Sequence 2, Appl
C 2	14.8	74.0	3636	4 US-09-388-774-2	Sequence 2, Appl
C 3	14.2	71.0	308	2 US-08-673-190A-4	Sequence 4, Appl
C 4	14.2	71.0	1235	3 US-08-672-850-5	Sequence 5, Appl
C 5	14.2	71.0	18318	1 US-08-414-926A-6	Sequence 6, Appl
C 6	14.2	71.0	18318	2 US-08-926-922-6	Sequence 6, Appl
C 7	14.2	71.0	18318	3 US-09-253-682-6	Sequence 6, Appl
C 8	14.2	71.0	18318	4 US-09-527-657-6	Sequence 6, Appl
C 9	13.8	69.0	1530	4 US-09-134-078-4	Sequence 6, Appl
C 10	13.8	69.0	1611	4 US-09-532-594B-3	Sequence 14, Appl
C 11	13.8	69.0	1872	4 US-09-532-594B-15	Sequence 15, Appl
C 12	13.8	69.0	1872	4 US-09-532-594B-15	Sequence 15, Appl
C 13	13.8	69.0	2475	4 US-09-624-693A-20	Sequence 20, Appl
C 14	13.8	69.0	3098	2 US-08-818-070-1	Sequence 1, Appl
C 15	13.8	69.0	3098	2 US-08-723-585-1	Sequence 1, Appl
C 16	13.8	69.0	4463	2 US-08-760-489-1	Sequence 1, Appl
C 17	13.8	69.0	4463	2 US-08-760-489-3	Sequence 1, Appl
C 18	13.8	69.0	4463	2 US-08-185-373-1	Sequence 1, Appl
C 19	13.8	69.0	4463	4 US-09-185-373-3	Sequence 1, Appl
C 20	13.8	69.0	4767	4 US-09-532-594B-1	Sequence 1, Appl
C 21	13.8	69.0	50000	4 US-09-146-053-4	Sequence 1, Appl
C 22	13.6	68.0	131	1 US-07-998-003A-92	Sequence 92, Appl
C 23	13.6	68.0	131	1 US-08-453-274B-92	Sequence 92, Appl
C 24	13.6	68.0	131	1 US-08-453-695A-92	Sequence 92, Appl
C 25	13.6	68.0	131	1 US-08-268-161A-92	Sequence 92, Appl
C 26	13.6	68.0	131	2 US-08-453-702A-92	Sequence 92, Appl
C 27	13.6	68.0	131	4 US-09-099-639-92	Sequence 92, Appl

28	13.6	68.0	131	5 PCT-US93-12588-92	Sequence 92, Appl
29	13.6	68.0	131	5 PCT-US95-08071-92	Sequence 92, Appl
C 30	13.6	68.0	1117	4 US-08-858-207A-40	Sequence 40, Appl
C 31	13.6	68.0	1503	4 US-09-172-952-6	Sequence 6, Appl
C 32	13.6	68.0	1509	2 US-08-481-337A-1	Sequence 1, Appl
C 33	13.6	68.0	1509	2 US-08-696-268B-1	Sequence 1, Appl
C 34	13.6	68.0	1509	5 PCT-US95-05467-1	Sequence 1, Appl
C 35	13.6	68.0	1506	5 PCT-US94-11328A-3	Sequence 1, Appl
C 36	13.6	68.0	1607	2 US-08-986-217-4	Sequence 4, Appl
C 37	13.6	68.0	1858	4 US-09-521-780-4	Sequence 4, Appl
C 38	13.6	68.0	1984	4 US-09-382-256-1	Sequence 1, Appl
C 39	13.6	68.0	1984	4 US-09-395-115-1	Sequence 1, Appl
C 40	13.6	68.0	1984	4 US-08-436-265-1	Sequence 1, Appl
C 41	13.6	68.0	1984	4 US-09-679-187-1	Sequence 1, Appl
C 42	13.6	68.0	3285	4 US-09-206-942-68	Sequence 68, Appl
C 43	13.6	68.0	3663	4 US-09-206-942-27	Sequence 27, Appl
C 44	13.6	68.0	3681	4 US-09-206-942-25	Sequence 25, Appl
C 45	13.6	68.0	3808	2 US-08-417-210A-79	Sequence 79, Appl

#### ALIGNMENTS

RESULT 1

US-09-074-579-2/c

; Sequence 2, Application US/09074579

; Patent No. 6001596

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Guebler, Karl J.

; APPLICANT: Patterson, Chandra

; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPsin-TYPE

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: IBM Compatible

; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2

; CURRENT APPLICATION DATA:

; FILING DATE: HEREWITH

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Cerrone, Michael C

; REGISTRATION NUMBER: 39,132

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELETYPE:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3636 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: UTRSN0702

; CLONE: 688183

; US-09-074-579-2

Query Match 74.0%; Score 14.8; DB 3; Length 3636;

Best Local Similarity 88.9%; Pred. No. 57;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ATATTAGCCCGCAGCTC 18  
DB 660 ATATTAGCTCCAGCTC 643

## RESULT 2

US-09-388-774-2/c  
; Sequence 2, Application US/09388774  
; Patent No. 6238991  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Patterson, Chandra  
; TITLE OF INVENTION: GROWTH-ASSOCIATED PREPSIN-TYPE  
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/388,774  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/074,579  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cerrone, Michael C  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PR-0505 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3636 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: UTRSN0702  
; CLONE: 688183  
; US-09-388-774-2

Query Match 74.0%; Score 14.8; DB 4; Length 3636;  
Best Local Similarity 88.9%; Pred. No. 57;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTAGCCCGCAGCTC 18  
DB 660 ATATTAGCTCCAGCTC 643

## RESULT 3

US-08-673-190A-4/c  
; Sequence 4, Application US/08673190A  
; Patent No. 5985668  
; GENERAL INFORMATION:  
; APPLICANT: Mattes, Ralf  
; APPLICANT: Klein, Kathrin

; APPLICANT: Stegmaier, Sabine  
; TITLE OF INVENTION: Sucrose Metabolism Mutants.  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/673,190A  
; FILING DATE: 27-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Forman, David S.  
; REGISTRATION NUMBER: 33,694  
; REFERENCE/DOCKET NUMBER: 06473.0001-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 308 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Enterobacter species  
; US-08-673-190A-4

Query Match 71.0%; Score 14.2; DB 2; Length 308;  
Best Local Similarity 84.2%; Pred. No. 94;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TATTAGCCCGCAGCTC 20  
DB 102 TATTAGCTCCAGCTC 84

## RESULT 4

US-08-672-850-5  
; Sequence 5, Application US/08672850  
; Patent No. 6140117  
; GENERAL INFORMATION:  
; APPLICANT: Miltbrandt, Jeffrey  
; APPLICANT: Arai, Toshiyuki  
; TITLE OF INVENTION: NINJURIN  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/672,850  
; FILING DATE: 24-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.

LOCATION: 1401..2384	OTHER INFORMATION: /product = "U1135"
FEATURE:	
NAME/KEY: CDS	LOCATION: 2478..3197
OTHER INFORMATION: /product = "U1136"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 3283..3789
OTHER INFORMATION: /product = "U1138"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 4355..4759
OTHER INFORMATION: /product = "U1139"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 4944..5285
OTHER INFORMATION: /product = "U1140"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 5558..6832
OTHER INFORMATION: /product = "U1141"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 6988..7825
OTHER INFORMATION: /product = "U1142"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 7813..8088
OTHER INFORMATION: /product = "U1143"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 8468..8995
OTHER INFORMATION: /product = "U1144"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 9327..9626
OTHER INFORMATION: /product = "U1145"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 9910..10260
OTHER INFORMATION: /product = "U1146"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 10328..10804
OTHER INFORMATION: /product = "U1147"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 11106..12053
OTHER INFORMATION: /product = "U1148"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 12133..12942
OTHER INFORMATION: /product = "U1132"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 13569..14210
OTHER INFORMATION: /product = "U1130"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 16216..16581
OTHER INFORMATION: /product = "U1149"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 1004..1528
OTHER INFORMATION: /product = "U1134"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 3063..3350
OTHER INFORMATION: /product = "U1137"	
FEATURE:	
NAME/KEY: CDS	LOCATION: 16337..18262

```
? OTHER INFORMATION: /product = "U1150"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 17752..18759
? OTHER INFORMATION: /product = "U1151"
? US-08-414-926A-6

Query Match          71.0%; Score 14.2; DB 1; Length 18318;
Best Local Similarity 84.2%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 AATATTACGCCACGCTCT 19
Db      15055 ATGTTACGCCCGCGCTCT 15073

RESULT 6
US-08-926-922-6
? Sequence 6, Application US/08926922
? Patent No. 5925751
? GENERAL INFORMATION:
? APPLICANT: Spaete, Richard
? APPLICANT: Cha, Tai-hn
? TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
? NUMBER OF SEQUENCES: 27
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Luann Cseir Attorney at Law
? STREET: 750 Arimo Avenue
? CITY: Oakland
? STATE: CA
? COUNTRY: USA
? ZIP: 94610

COMPUTER READABLE FORM:
MEDION TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/926 922
FILING DATE: September 10, 1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Cseir, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AYTR 11A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-834-1448
TELEFAX: 510-839-7810
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 18318 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Human CMV
STRAIN: Toledo
FEATURE:
NAME/KEY: CDS
LOCATION: 511..1281
OTHER INFORMATION: /product = "U1133"
FEATURE:
NAME/KEY: CDS
LOCATION: 1401..2384
OTHER INFORMATION: /product = "U1135"
FEATURE:
NAME/KEY: CDS
LOCATION: 2478..3197
OTHER INFORMATION: /product = "U1136"
FEATURE:

? NAME/KEY: CDS
? LOCATION: 3283..3789
? OTHER INFORMATION: /product = "U1138"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 4355..4759
? OTHER INFORMATION: /product = "U1139"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 4944..5285
? OTHER INFORMATION: /product = "U1140"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 5558..6832
? OTHER INFORMATION: /product = "U1141"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 6908..7825
? OTHER INFORMATION: /product = "U1142"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 7813..8088
? OTHER INFORMATION: /product = "U1143"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 8468..8935
? OTHER INFORMATION: /product = "U1144"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 9327..9626
? OTHER INFORMATION: /product = "U1145"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 9910..10260
? OTHER INFORMATION: /product = "U1146"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 10328..10804
? OTHER INFORMATION: /product = "U1147"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 11106..12053
? OTHER INFORMATION: /product = "U1148"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 12133..12942
? OTHER INFORMATION: /product = "U1132"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 13569..14210
? OTHER INFORMATION: /product = "U1130"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 16216..16581
? OTHER INFORMATION: /product = "U1149"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1004..1528
? OTHER INFORMATION: /product = "U1134"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3063..3350
? OTHER INFORMATION: /product = "U1137"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 16337..18262
? OTHER INFORMATION: /product = "U1150"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 17752..18759
? OTHER INFORMATION: /product = "U1151"
? FEATURE:
? US-08-926-922-6
```

Query Match 71.0%, Score 14.2; DB 2; Length 18318;  
Best Local Similarity 84.2%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTACGCCGCTCT 19

DB 15055 ATGTACGCCGCTCT 15073

## RESULT 7

US-09-253-682-6

Sequence 6, Application US/09253682

Patent No. 6040170

GENERAL INFORMATION:

APPLICANT: Spaete, Richard

APPLICANT: Cha, Tai-An

TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Luann Cseerr Attorney at Law

STREET: 750 Arimo Avenue

CITY: Oakland

STATE: CA

COUNTRY: USA

ZIP: 94610

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/253,682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/926,922

FILING DATE: September 10, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Cseerr, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR 11A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-834-1448

TELEFAX: 510-839-7810

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 18318 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORGANISM: Human CMV

STRAIN: Toledo

FEATURE:

NAME/KEY: CDS

LOCATION: 511..1281

OTHER INFORMATION: /product = "UL133"

FEATURE:

NAME/KEY: CDS

LOCATION: 1401..2384

OTHER INFORMATION: /product = "UL135"

FEATURE:

NAME/KEY: CDS

LOCATION: 2478..3197

OTHER INFORMATION: /product = "UL136"

FEATURE:

NAME/KEY: CDS

LOCATION: 3283..3789

OTHER INFORMATION: /product = "UL138"

FEATURE:

NAME/KEY: CDS  
LOCATION: 4355..4759  
OTHER INFORMATION: /product = "UL139"

FEATURE:

NAME/KEY: CDS

LOCATION: 4944..5285

OTHER INFORMATION: /product = "UL140"

FEATURE:

NAME/KEY: CDS

LOCATION: 5558..6832

OTHER INFORMATION: /product = "UL141"

FEATURE:

NAME/KEY: CDS

LOCATION: 6908..7825

OTHER INFORMATION: /product = "UL142"

FEATURE:

NAME/KEY: CDS

LOCATION: 7813..8088

OTHER INFORMATION: /product = "UL143"

FEATURE:

NAME/KEY: CDS

LOCATION: 8468..8995

OTHER INFORMATION: /product = "UL144"

FEATURE:

NAME/KEY: CDS

LOCATION: 9327..9626

OTHER INFORMATION: /product = "UL145"

FEATURE:

NAME/KEY: CDS

LOCATION: 9910..10260

OTHER INFORMATION: /product = "UL146"

FEATURE:

NAME/KEY: CDS

LOCATION: 10328..10804

OTHER INFORMATION: /product = "UL147"

FEATURE:

NAME/KEY: CDS

LOCATION: 11106..12053

OTHER INFORMATION: /product = "UL148"

FEATURE:

NAME/KEY: CDS

LOCATION: 12133..12942

OTHER INFORMATION: /product = "UL149"

FEATURE:

NAME/KEY: CDS

LOCATION: 13569..14210

OTHER INFORMATION: /product = "UL150"

FEATURE:

NAME/KEY: CDS

LOCATION: 16216..16581

OTHER INFORMATION: /product = "UL151"

FEATURE:

NAME/KEY: CDS

LOCATION: 1004..1528

OTHER INFORMATION: /product = "UL152"

FEATURE:

NAME/KEY: CDS

LOCATION: 16337..18262

OTHER INFORMATION: /product = "UL153"

FEATURE:

NAME/KEY: CDS

LOCATION: 17752..18759

OTHER INFORMATION: /product = "UL154"

FEATURE:

NAME/KEY: CDS

LOCATION: 17752..18759

OTHER INFORMATION: /product = "UL155"

FEATURE:

Query Match 71.0%; Score 14.2; DB 3; Length 18318;

Best Local Similarity 84.2%; Pred. No. 1.4e+02;

Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTAACGCCGCGCTCT 19  
|||  
Db 15055 ATGTTAAGCCCGCGCTCT 15073

## RESULT 8

US-09-527-657-6

Sequence 6, Application US/09527657

Patent No. 6291236

## GENERAL INFORMATION:

APPLICANT: Spaete, Richard

Cha, Tai-nn

TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Luanm Cserr Attorney at Law

STREET: 750 Arimo Avenue

CITY: Oakland

STATE: CA

COUNTRY: USA

ZIP: 94610

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/09/527,657

FILING DATE: 17-Mar-2000

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/926,922

FILING DATE: September 10, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Cserr, Luanm

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR 11A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 510-834-1448

TELEFAX: 510-839-7810

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 18318 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Human CMV

STRAIN: Toledo

## FEATURE:

NAME/KEY: CDS

LOCATION: 511..1281

OTHER INFORMATION: /product = "UL133"

## FEATURE:

NAME/KEY: CDS

LOCATION: 1401..2384

OTHER INFORMATION: /product = "UL135"

## FEATURE:

NAME/KEY: CDS

LOCATION: 2478..3197

OTHER INFORMATION: /product = "UL136"

## FEATURE:

NAME/KEY: CDS

LOCATION: 3283..3789

OTHER INFORMATION: /product = "UL138"

## FEATURE:

NAME/KEY: CDS

LOCATION: 4355..4759

OTHER INFORMATION: /product = "UL139"

## FEATURE:

NAME/KEY: CDS

LOCATION: 4355..4759

OTHER INFORMATION: /product = "UL139"

## FEATURE:

NAME/KEY: CDS

LOCATION: 4355..4759

OTHER INFORMATION: /product = "UL139"

## FEATURE:

NAME/KEY: CDS  
LOCATION: 4944..5285  
OTHER INFORMATION: /product = "UL140"

FEATURE:

NAME/KEY: CDS  
LOCATION: 5558..6832  
OTHER INFORMATION: /product = "UL141"

FEATURE:

NAME/KEY: CDS  
LOCATION: 6908..7825  
OTHER INFORMATION: /product = "UL142"

FEATURE:

NAME/KEY: CDS  
LOCATION: 7813..8088  
OTHER INFORMATION: /product = "UL143"

FEATURE:

NAME/KEY: CDS  
LOCATION: 8468..8995  
OTHER INFORMATION: /product = "UL144"

FEATURE:

NAME/KEY: CDS  
LOCATION: 9327..9626  
OTHER INFORMATION: /product = "UL145"

FEATURE:

NAME/KEY: CDS  
LOCATION: 9910..10260  
OTHER INFORMATION: /product = "UL146"

FEATURE:

NAME/KEY: CDS  
LOCATION: 10328..10804  
OTHER INFORMATION: /product = "UL147"

FEATURE:

NAME/KEY: CDS  
LOCATION: 11106..12053  
OTHER INFORMATION: /product = "UL148"

FEATURE:

NAME/KEY: CDS  
LOCATION: 12133..12942  
OTHER INFORMATION: /product = "UL132"

FEATURE:

NAME/KEY: CDS  
LOCATION: 13569..14210  
OTHER INFORMATION: /product = "UL130"

FEATURE:

NAME/KEY: CDS  
LOCATION: 16216..16581  
OTHER INFORMATION: /product = "UL149"

FEATURE:

NAME/KEY: CDS  
LOCATION: 1004..1528  
OTHER INFORMATION: /product = "UL134"

FEATURE:

NAME/KEY: CDS  
LOCATION: 3063..3350  
OTHER INFORMATION: /product = "UL137"

FEATURE:

NAME/KEY: CDS  
LOCATION: 16337..18262  
OTHER INFORMATION: /product = "UL150"

FEATURE:

NAME/KEY: CDS  
LOCATION: 17752..18759  
OTHER INFORMATION: /product = "UL151"

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-527-657-6

Query Match 71.0%; Score 14.2; DB 4; Length 18318;  
Best Local Similarity 84.2%; Pred. No. 1,4e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTAACGCCGCGCTCT 19  
|||  
Db 15055 ATGTTAAGCCCGCGCTCT 15073

```

RESULT 9
US-09-134-078-4
; Sequence 4, Application US/09134078
; Patent No. 6368844
; GENERAL INFORMATION:
; APPLICANT: Bylina, Edward J.
; TITLE OF INVENTION: GYCOSIDASE ENZYMES
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESS: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,078
; FILING DATE: 13-AUG-1998
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/949,026
; FILING DATE: 10-OCT-1997
; APPLICATION NUMBER: 60/056,916
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/024002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858/677-1456
; TELEFAX: 858/677-1465
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1530 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..1527
; US-09-134-078-4

Query Match          69.0%; Score 13.8; DB 4; Length 1530;
Best Local Similarity 88.2%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATTACGCCACGCTCTC 19
Db 727 ATTACGCCACGCTCTC 743

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1611
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAY4 Rep 68 gene
; US-09-532-594B-14

Query Match          69.0%; Score 13.8; DB 4; Length 1611;
Best Local Similarity 88.2%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TTACGCCACGCTCTC 20
Db 302 TTACGCCACGCTCTC 286

RESULT 11
US-09-532-594B-3/C
; Sequence 3, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Safet, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAY4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: CDS
; LOCATION: (1)...(1872)
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAY4 Rep gene (full length)
; US-09-532-594B-3

Query Match          69.0%; Score 13.8; DB 4; Length 1872;
Best Local Similarity 88.2%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TTACGCCACGCTCTC 20
Db 302 TTACGCCACGCTCTC 286

RESULT 12
US-09-532-594B-15/C
; Sequence 15, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Safet, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAY4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0

```

```
; SEQ ID NO 15
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 Rep 78 gene
US-09-532-594B-15

Query Match
Best Local Similarity 69.0%; Score 13.8; DB 4; Length 1872;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TTACGCCACGCTCTC 20
DB 302 TTGACGCCACGCTCTC 286

RESULT 13
US-09-624-693A-20
; Sequence 20, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 2475
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_difference
; LOCATION: (13, 34, 46, 49, 51, 57, 59, 68, 69, 73, 75 - 77, 79, 82, 84,
; OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
; OTHER INFORMATION: Description of Artificial Sequence: Consensus
; Patent No. 6355468
US-09-624-693A-20

Query Match
Best Local Similarity 69.0%; Score 13.8; DB 4; Length 2475;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ATTAAGCCCCAGCTCTC 20
DB 1162 ATTACGCCACGCTCTC 1179

RESULT 14
US-08-818-070-1
; Sequence 1, Application US/08818070
; Patent No. 5831019
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy
; APPLICANT: Bolwig, Gert M.
; TITLE OF INVENTION: ASSOCIATIVE LEARNING AND THE linotte
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, PC
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
```

```
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,070
; FILING DATE: 14-MAR-1997
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/723,585
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL95-03PA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3098 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 89...2833
US-08-818-070-1

Query Match
Best Local Similarity 69.0%; Score 13.8; DB 2; Length 3098;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TTACGCCACGCTCTC 20
DB 803 TTAGGCCACGCTCTC 819

RESULT 15
US-08-723-585-1
; Sequence 1, Application US/08723585
; Patent No. 5831057
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy
; APPLICANT: Bolwig, Gert M.
; TITLE OF INVENTION: ASSOCIATIVE LEARNING AND THE linotte
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, PC
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,585
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: CSHL95-03PA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
```

TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3098 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 89..2833  
US-08-723-585-1

Query Match 69.0%; Score 13.8; DB 2; Length 3098;  
Best Local Similarity 88.2%; Pred. No. 1.9e+02;  
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 4 TTACGCCACGCTCTC 20  
||| |||||  
DB 803 TTAGCCCAACGCTCTC 819

Search completed: May 12, 2003, 02:38:44  
Job time : 21.7857 secs

GenoCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 02:35:34 ; Search time 18.7143 Seconds  
(without alignments)  
1328.082 Million cell updates/sec

Title: US-09-980-277-3  
Perfect score: 20  
Sequence: 1 atattacgcgcacgcctc 20

Scoring table: IDENTIFY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 783654 seqs, 621352466 residues  
Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCF\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCFUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.4	77.0	2535	10 US-09-815-242-7400	Sequence 7400, Ap
2	15.4	77.0	2577	10 US-09-815-242-7299	Sequence 7299, Ap
3	15.2	76.0	382	10 US-09-878-574-857	Sequence 857, Ap
4	15.2	76.0	385	9 US-09-918-995-37010	Sequence 37010, A
5	15.2	76.0	1529	10 US-09-820-893-20	Sequence 20, Appl
6	15.2	76.0	1796	10 US-09-820-893-43	Sequence 43, Appl
7	15.2	76.0	2033	9 US-09-981-876-64	Sequence 64, Appl
8	15.2	76.0	2033	9 US-09-148-545-64	Sequence 64, Appl
9	15.2	76.0	2136	9 US-09-981-876-120	Sequence 120, Appl
10	15.2	76.0	2136	9 US-09-148-545-120	Sequence 120, Appl
11	15.2	76.0	2136	10 US-09-820-893-44	Sequence 44, Appl
12	15.2	76.0	2196	10 US-09-917-800A-1469	Sequence 1469, Ap
13	15.2	76.0	6061	9 US-10-239-676-40	Sequence 40, Appl
14	15.2	74.0	873	9 US-09-738-626-1562	Sequence 1562, Ap
15	14.8	74.0	2331	9 US-09-944-413-54	Sequence 54, Appl
16	14.8	74.0	2331	9 US-09-944-403-54	Sequence 54, Appl
17	14.8	74.0	2331	9 US-09-944-896-54	Sequence 54, Appl
18	14.8	74.0	2331	9 US-09-944-944-54	Sequence 54, Appl
19	14.8	74.0	2331	9 US-09-944-907-54	Sequence 54, Appl

C 20	14.8	74.0	2331	9	US-09-944-929-54	Sequence 54, Appl
C 21	14.8	74.0	2331	9	US-10-174-590-59	Sequence 59, Appl
C 22	14.8	74.0	2331	9	US-10-176-758-59	Sequence 59, Appl
C 23	14.8	74.0	2331	9	US-10-175-737-59	Sequence 59, Appl
C 24	14.8	74.0	2331	9	US-10-173-706-59	Sequence 59, Appl
C 25	14.8	74.0	2331	9	US-10-175-738-59	Sequence 59, Appl
C 26	14.8	74.0	2331	9	US-10-175-752-59	Sequence 59, Appl
C 27	14.8	74.0	2331	9	US-10-176-482-59	Sequence 59, Appl
C 28	14.8	74.0	2331	9	US-10-176-757-59	Sequence 59, Appl
C 29	14.8	74.0	2331	9	US-10-176-913-59	Sequence 59, Appl
C 30	14.8	74.0	2331	9	US-10-180-552-59	Sequence 59, Appl
C 31	14.8	74.0	2331	9	US-10-180-557-59	Sequence 59, Appl
C 32	14.8	74.0	2331	9	US-10-173-700-59	Sequence 59, Appl
C 33	14.8	74.0	2331	9	US-10-174-572-59	Sequence 59, Appl
C 34	14.8	74.0	2331	9	US-10-174-579-59	Sequence 59, Appl
C 35	14.8	74.0	2331	9	US-10-174-582-59	Sequence 59, Appl
C 36	14.8	74.0	2331	9	US-10-174-588-59	Sequence 59, Appl
C 37	14.8	74.0	2331	9	US-10-175-739-59	Sequence 59, Appl
C 38	14.8	74.0	2331	9	US-10-175-740-59	Sequence 59, Appl
C 39	14.8	74.0	2331	9	US-10-175-743-59	Sequence 59, Appl
C 40	14.8	74.0	2331	9	US-10-176-488-59	Sequence 59, Appl
C 41	14.8	74.0	2331	9	US-10-176-492-59	Sequence 59, Appl
C 42	14.8	74.0	2331	9	US-10-176-747-59	Sequence 59, Appl
C 43	14.8	74.0	2331	9	US-10-176-750-59	Sequence 59, Appl
C 44	14.8	74.0	2331	9	US-10-176-985-59	Sequence 59, Appl
C 45	14.8	74.0	2331	9	US-10-176-987-59	Sequence 59, Appl

## ALIGNMENTS

RESULT 1  
US-09-815-242-7400/c  
Sequence 7400, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlson, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA, 011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/131,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 7400  
LENGTH: 2535  
TYPE: DNA  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(2535)  
US-09-815-242-7400

```
Query Match          77.0%; Score 15.4; DB 10; Length 2535;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 ATTAAGCCCGACGCTCT 19
      ||| ||||| ||||| |||
Db      1638 ATTCAGCCCGACGCTCT 1622

RESULT 2
US-09-815-242-7299/c
; Sequence 7299, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trivick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7299
; LENGTH: 2577
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2577)
US-09-815-242-7299

Query Match          77.0%; Score 15.4; DB 10; Length 2577;
Best Local Similarity 94.1%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 ATTAAGCCCGACGCTCT 19
      ||| ||||| ||||| |||
Db      1680 ATTCAGCCCGACGCTCT 1664

RESULT 3
US-09-878-574-857
; Sequence 857, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
```

```
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 857
; LENGTH: 382
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-045-Q1-B1-E8
US-09-878-574-857

Query Match          76.0%; Score 15.2; DB 10; Length 382;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ATATTAGCCCGACGCTCTC 20
      ||||| ||||| ||||| |||
Db      58 ATATTAGCCCGACGCTCTCTC 77

RESULT 4
US-09-918-995-37010/c
; Sequence 37010, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37010
; LENGTH: 395
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-37010

Query Match          76.0%; Score 15.2; DB 9; Length 395;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ATATTAGCCCGACGCTCTC 20
      ||||| ||||| ||||| |||
Db      30 ATATTAGCTCAGCTCTCTC 11

RESULT 5
US-09-820-893-20/c
; Sequence 20, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 31 Human Secreted Proteins
; FILE REFERENCE: P2033P1
; CURRENT APPLICATION NUMBER: US/09/820,893
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/531,119
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 60/102,895
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
```

```
; LOCATION: (1505)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-820-893-20

Query Match          76.0%; Score 15.2; DB 10; Length 1529;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTACGCCACGCTCTC 20
   |||||  |||  |||  |||
Db 144 ATATTACGTCATGTTCTC 125

RESULT 6
US-09-820-893-43/c
; Sequence 43, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 31 Human Secreted Proteins
; FILE REFERENCE: P2033P1
; CURRENT APPLICATION NUMBER: US/09/820,893
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/531,119
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 60/102,895
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1796
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (417)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-820-893-43

Query Match          76.0%; Score 15.2; DB 10; Length 1796;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTACGCCACGCTCTC 20
   |||||  |||  |||  |||
Db 66 ATATTACGTCATGTTCTC 47

RESULT 7
US-09-981-876-64/c
; Sequence 64, Application US/09981876
; Patent No. US20020164669A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/981,876
; CURRENT FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: 09/148,545
; PRIOR FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,333
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/038,621
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,161
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,626
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,334
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,336

; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,163
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/047,615
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,600
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,597
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,502
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,633
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,583
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,617
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,618
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,503
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,592
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,581
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,584
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,500
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,587
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,492
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,598
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,613
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,582
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,596
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,612
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,568
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,569
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,311
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,671
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,674
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,669
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,312
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,313
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,672
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,315
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/048,974
; PRIOR FILING DATE: 1997-06-06
```

1	PRIOR APPLICATION NUMBER: 60/056, 886
2	PRIOR FILING DATE: 1997-08-22
3	PRIOR APPLICATION NUMBER: 60/056, 877
4	PRIOR FILING DATE: 1997-08-22
5	PRIOR APPLICATION NUMBER: 60/056, 889
6	PRIOR FILING DATE: 1997-08-22
7	PRIOR APPLICATION NUMBER: 60/056, 893
8	PRIOR FILING DATE: 1997-08-22
9	PRIOR APPLICATION NUMBER: 60/056, 630
10	PRIOR FILING DATE: 1997-08-22
11	PRIOR APPLICATION NUMBER: 60/056, 878
12	PRIOR FILING DATE: 1997-08-22
13	PRIOR APPLICATION NUMBER: 60/056, 662
14	PRIOR FILING DATE: 1997-08-22
15	PRIOR APPLICATION NUMBER: 60/056, 872
16	PRIOR FILING DATE: 1997-08-22
17	PRIOR APPLICATION NUMBER: 60/056, 882
18	PRIOR FILING DATE: 1997-08-22
19	PRIOR APPLICATION NUMBER: 60/056, 637
20	PRIOR FILING DATE: 1997-08-22
21	PRIOR APPLICATION NUMBER: 60/056, 880
22	PRIOR FILING DATE: 1997-08-22
23	PRIOR APPLICATION NUMBER: 60/056, 888
24	PRIOR FILING DATE: 1997-08-22
25	PRIOR APPLICATION NUMBER: 60/056, 894
26	PRIOR FILING DATE: 1997-08-22
27	PRIOR APPLICATION NUMBER: 60/056, 911
28	PRIOR FILING DATE: 1997-08-22
29	PRIOR APPLICATION NUMBER: 60/056, 636
30	PRIOR FILING DATE: 1997-08-22
31	PRIOR APPLICATION NUMBER: 60/056, 874
32	PRIOR FILING DATE: 1997-08-22
33	PRIOR APPLICATION NUMBER: 60/056, 910
34	PRIOR FILING DATE: 1997-08-22
35	PRIOR APPLICATION NUMBER: 60/056, 866
36	PRIOR FILING DATE: 1997-08-22
37	PRIOR APPLICATION NUMBER: 60/056, 633
38	PRIOR FILING DATE: 1997-08-22
39	PRIOR APPLICATION NUMBER: 60/056, 845
40	PRIOR FILING DATE: 1997-08-22
41	PRIOR APPLICATION NUMBER: 60/056, 892
42	PRIOR FILING DATE: 1997-08-22
43	PRIOR APPLICATION NUMBER: 60/047, 595
44	PRIOR FILING DATE: 1997-05-23
45	PRIOR APPLICATION NUMBER: 60/057, 761
46	PRIOR FILING DATE: 05-SEP-1997
47	PRIOR APPLICATION NUMBER: 60/047, 599
48	PRIOR FILING DATE: 1997-05-23
49	PRIOR APPLICATION NUMBER: 60/047, 586
50	PRIOR FILING DATE: 1997-05-23
51	PRIOR APPLICATION NUMBER: 60/047, 585
52	PRIOR FILING DATE: 1997-05-23
53	PRIOR APPLICATION NUMBER: 60/047, 584
54	PRIOR FILING DATE: 1997-05-23
55	PRIOR APPLICATION NUMBER: 60/047, 590
56	PRIOR FILING DATE: 1997-05-23
57	PRIOR APPLICATION NUMBER: 60/047, 594
58	PRIOR FILING DATE: 1997-05-23
59	PRIOR APPLICATION NUMBER: 60/047, 589
60	PRIOR FILING DATE: 1997-05-23
61	PRIOR APPLICATION NUMBER: 60/043, 578
62	PRIOR FILING DATE: 1997-04-11
63	PRIOR APPLICATION NUMBER: 60/043, 576
64	PRIOR FILING DATE: 1997-04-11
65	PRIOR APPLICATION NUMBER: 60/047, 501

```

PRIORITY FILLING DATE: 1997-05-23
PRIORITY APPLICATION NUMBER: 60/043,670
PRIORITY FILLING DATE: 1997-04-11
PRIORITY APPLICATION NUMBER: 60/056,632
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/056,664
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/056,876
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/056,881
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/056,909
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/056,875
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/056,862
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/056,887
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/056,908
PRIORITY FILLING DATE: 1997-08-22
PRIORITY APPLICATION NUMBER: 60/048,964
PRIORITY FILLING DATE: 1997-06-06
PRIORITY APPLICATION NUMBER: 60/057,650
PRIORITY FILLING DATE: 1997-09-05
PRIORITY APPLICATION NUMBER: 60/056,884
PRIORITY FILLING DATE: 1997-08-22
NUMBER OF SEQ ID NOS: 280
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 64
LENGTH: 2033

Query Match          76.0%; Score 15.2; DB 9; Length 2033;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ATATTACGCCACGCTCTC 20
        |||||  |||  ||  |||
DB      43 ATATTACGCCACGCTCTC 24

RESULT 8
US-09-148-545-64/c
Sequence 64, Application US/09148545
Publication No. US20030027132A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 70 Human Secreted Proteins
FILE REFERENCE: P2001P1
CURRENT APPLICATION NUMBER: US/09/148,545
EARLIER APPLICATION NUMBER: PCT/US98/04482
EARLIER FILLING DATE: 1998-09-04
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILLING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILLING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILLING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,161
EARLIER FILLING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILLING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILLING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILLING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILLING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILLING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILLING DATE: 1997-05-23

```

[illegible]

```

; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 64
; LENGTH: 2033

Query Match          76.0%; Score 15.2; DB 9; Length 2033;
Best Local Similarity 85.0%; Pred. No. 1,3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTAGCCGACGCTC 20
    |||||  |||  |||
DB 43 ATATTAGCTCATGTTCTC 24

RESULT 9
US-09-981-876-120/c
; Sequence 120, Application US/09981876
; Patent No. US2002016469A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/981,876
; CURRENT FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: 09/148,545
; PRIOR FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,333
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/038,621
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,161
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,626
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,334
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,336
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,163
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/047,615
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,600
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,597
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,502
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,633
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,583

; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,617
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,618
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,503
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,592
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,581
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,584
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,500
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,587
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,492
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,598
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,613
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,582
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,596
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,612
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,568
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,569
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,311
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,671
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,674
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,669
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,312
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,313
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,672
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/043,315
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: 60/048,974
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/056,886
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,877
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,889
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,893
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,630
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,878
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 60/056,662
; PRIOR FILING DATE: 1997-08-22
```

;; PRIOR APPLICATION NUMBER: 60/056,872  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,882  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,637  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,903  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,888  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,879  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,880  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,894  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,911  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,636  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,874  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,910  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,864  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,631  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,845  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,892  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/047,595  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/057,761  
;; PRIOR FILING DATE: 05-Sep-1997  
;; PRIOR APPLICATION NUMBER: 60/047,599  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/047,588  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/047,585  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/047,586  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/047,590  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/047,594  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/047,589  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/047,593  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/047,614  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/043,578  
;; PRIOR FILING DATE: 1997-04-11  
;; PRIOR APPLICATION NUMBER: 60/043,576  
;; PRIOR FILING DATE: 1997-04-11  
;; PRIOR APPLICATION NUMBER: 60/047,501  
;; PRIOR FILING DATE: 1997-05-23  
;; PRIOR APPLICATION NUMBER: 60/043,670  
;; PRIOR FILING DATE: 1997-04-11  
;; PRIOR APPLICATION NUMBER: 60/056,632  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,664  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,876  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,881  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,909  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,875

;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,862  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,887  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/056,908  
;; PRIOR FILING DATE: 1997-08-22  
;; PRIOR APPLICATION NUMBER: 60/048,964  
;; PRIOR FILING DATE: 1997-06-06  
;; PRIOR APPLICATION NUMBER: 60/057,650  
;; PRIOR FILING DATE: 1997-09-05  
;; PRIOR APPLICATION NUMBER: 60/056,884  
;; PRIOR FILING DATE: 1997-08-22  
;; NUMBER OF SEQ ID NOS: 280  
;; SOFTWARE: Patent In Ver. 2.0  
;; SEQ ID NO 120  
;; LENGTH: 2136  
  
Query Match 76.0%; Score 15.2; DB 9; Length 2136;  
Best Local Similarity 85.0%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 ATATTAGGCCACGCTCTC 20  
Db 114 ATATTAGGCCACGCTCTC 95  
  
RESULT 10  
US-09-148-545-120/c  
; Sequence 120, Application US/09148545  
; Publication No. US20030027132A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 70 Human Secreted Proteins  
; FILE REFERENCE: P2001P1  
; CURRENT APPLICATION NUMBER: US/09/148,545  
; CURRENT FILING DATE: 1998-09-04  
; EARLIER APPLICATION NUMBER: PCT/US98/04482  
; EARLIER FILING DATE: 1998-03-06  
; EARLIER APPLICATION NUMBER: 60/040,162  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,333  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/038,621  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,161  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,626  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,334  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,336  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,163  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/047,615  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,600  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,597  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,502  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,633  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,583  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,617  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,618  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,503  
; EARLIER FILING DATE: 1997-05-23

[illegible]

```
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 280
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 120
; LENGTH: 2136

Query Match          76.0%; Score 15.2; DB 9; Length 2136;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTACGCCACGCTCTC 20
Db 114 ATATTACGTCACGTCTC 95

RESULT 11
US-09-820-893-44/c
; Sequence 44, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 31 Human Secreted Proteins
; FILE REFERENCE: P2033P1
; CURRENT APPLICATION NUMBER: US/09/820,893
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/531,119
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 60/102,895
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 44
; LENGTH: 2136
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-820-893-44

Query Match          76.0%; Score 15.2; DB 10; Length 2136;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTACGCCACGCTCTC 20
Db 114 ATATTACGTCACGTCTC 95

RESULT 12
US-09-917-800A-1469/c
; Sequence 1469, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendlick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Blashoff, Michael
; APPLICANT: Gese Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
```

```
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO: 1469
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 U63923
US-09-917-800A-1469

Query Match          76.0%; Score 15.2; DB 10; Length 2196;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTACGCCACGCTCTC 20
Db 1054 ATCTTACGCCACGCTCTC 1035

RESULT 13
US-10-239-676-40/c
; Sequence 40, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERBERCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO: 40
; LENGTH: 6061
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-40

Query Match          76.0%; Score 15.2; DB 9; Length 6061;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATATTACGCCACGCTCTC 20
Db 2190 ATATTACGTCACGTCTC 2171

RESULT 14
US-09-738-626-1562
; Sequence 1562, Application US/09738626
```

```

; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1562
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1562

Query Match          74.0%; Score 14.8; DB 9; Length 873;
Best Local Similarity 88.9%; Fred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 TATTACGCCACGCTCT 19
        ||| ||||| ||||| ||
Db       579 TATTACGCCACGCTCT 596

RESULT 15
US-09-944-413-54/c
; Sequence 54, Application US/0944413
; Patent No. US20020156004A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Baton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Ellvaroff, Ellen
; APPLICANT: Geriltsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gutney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kijavlin, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Thomas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,413
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,335
; PRIOR FILING DATE: December 11, 1997
```

```

; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 15, 1997
; PRIOR APPLICATION NUMBER: 60/069,870
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/068,017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074,092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075,945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146,222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020156004A1, December 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020156004A1, December 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
; PRIOR FILING DATE: July 28, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 54
; LENGTH: 2331
; TYPE: DNA
; ORGANISM: Homo Sapien
```

US-09-944-413-54

Query Match 74.0%; Score 14.8; DB 9; Length 2331;  
Best local Similarity 88.9%; Pred. No. 2.1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTACGCCACGCTC 18  
|||||  
Db 643 ATATTACGCTCCACGCTC 626

Search completed: May 12, 2003, 04:49:52  
Job time : 20.7143 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 01:04:28 ; Search time 244.143 Seconds  
(without alignments)  
1326.722 Million cell updates/sec

Title: US-09-980-277-3

Perfect score: 20

Sequence: 1 attataacgccacgcctctc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: em\_estha.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estinu.\*  
5: em\_estov.\*  
6: em\_estopl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_estcl.\*  
10: gb\_estcl2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	687	17	BH210086
2	17.4	87.0	17	17	AQ464286 HS_5101_A
3	16.8	84.0	319	17	AZ216980 Sheared D
4	16.8	84.0	706	12	BG429009 602501922
5	16.8	84.0	804	10	AV755514 AV755514
6	16.4	82.0	219	9	AA698952 HL06045.5

Result No.	Score	Query Match	Length	ID	Description
7	16.4	82.0	294	17	BH809948
8	16.4	82.0	307	9	AA698954
9	16.4	82.0	573	13	B1962036
10	16.4	82.0	679	13	BM133610
11	16.4	82.0	700	9	AA940936
12	16.4	82.0	872	9	AA950399
13	16.4	82.0	1503	17	AQ752039
14	16.4	82.0	940	17	CNS01MAP
15	16.4	82.0	1161	14	B0067992
16	16.4	82.0	1382	12	B639617
17	16.4	82.0	400	9	AI210561
18	16.4	82.0	401	14	Z14884
19	16.4	82.0	441	17	AQ908431
20	16.4	82.0	459	17	AQ707815
21	16.4	82.0	487	17	AQ406537
22	16.4	82.0	491	17	AZ183441
23	16.4	82.0	514	9	AU208418
24	16.4	82.0	605	13	B0119342
25	16.4	82.0	621	17	AZ226253
26	16.4	82.0	644	12	B6851049
27	16.4	82.0	664	17	AZ870610
28	16.4	82.0	755	13	B1524144
29	16.4	82.0	838	9	AI648350
30	16.4	82.0	897	12	B6287095
31	16.4	82.0	931	12	B7141420
32	16.4	82.0	947	17	AG099094
33	16.4	82.0	973	17	CNS020AT
34	16.4	82.0	1014	17	AG111100
35	16.4	82.0	1028	17	AG097971
36	16.4	82.0	1112	12	B6686423
37	16.4	82.0	1179	12	B0953557
38	16.4	82.0	1650	14	BQ706996
39	16.4	82.0	235	17	AZ910305
40	16.4	82.0	320	12	B6339308
41	16.4	82.0	394	17	AO084176
42	16.4	82.0	421	10	AV789035
43	16.4	82.0	441	12	B6042015
44	16.4	82.0	441	12	B6044104
45	16.4	82.0	667	17	AG101516

## ALIGNMENTS

RESULT 1  
BH210086/c  
LOCUS BH210086 687 bp DNA linear GSS 24-OCT-2001  
DEFINITION Sml-56J20.TF Sml Schistosoma mansoni genomic clone Sml-56J20, DNA  
ACCESSION BH210086  
VERSION BH210086.1 GI:16389299  
KEYWORDS GSS.  
SOURCE Schistosoma mansoni.  
ORGANISM Schistosoma mansoni  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
REFERENCE Shetty,V., Simpson,A., Malek,J., Koo,H., Loverde,P.T. and El-Sayed  
,N.M.  
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)  
JOURNAL The Institute for Genomic Research  
COMMENT Unpublished (2001)  
Other GSSs: Sml-56J20..TR  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@igf.org  
lo.edu)  
Seq primer: M13 For  
Class: BAC ends.

## FEATURES

## source

Location/Qualifiers

1. .687  
/organism="Schistosoma mansoni"  
/strain="Puerto Rico"  
/db\_xref="taxon:6183"  
/clone="Sml-56J20"

/clone\_id="Sml"

/note="Vector: pBelobAC11; Site\_1: Hln diti; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hln diti. High molecular weight fragments were ligated in pBelobAC11 digested with Hln diti. The average insert size is 100 kb. Total clone coverage: approx. 7.95 x the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

BASE COUNT 218 a 101 c 152 g 216 t  
ORIGIN

## Query Match

Best Local Similarity 100.0%; Score 20; DB 17; Length 687;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATATTAGGCCACGCTCTC 20

Db 336 ATATTAGGCCACGCTCTC 317

## RESULT 2

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## /organism="Homo sapiens"

## /strain="taxon:9606"

## /db\_xref="taxon:9606"

## /clone="Plate=677 Col=15 Row=C"

## /clone\_lib="RPCT-11 Human Male BAC Library"

## BASE COUNT

## ORIGIN

## Query Match

## Best Local Similarity

## Score 20; DB 17; Length 687;

## Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY 1 ATATTAGGCCACGCTCTC 20

## Db 336 ATATTAGGCCACGCTCTC 317

## RESULT 2

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## /organism="Homo sapiens"

## /strain="taxon:9606"

## /db\_xref="taxon:9606"

## /clone="Plate=677 Col=15 Row=C"

## /clone\_lib="RPCT-11 Human Male BAC Library"

## BASE COUNT

## ORIGIN

## Query Match

## Best Local Similarity

/sex="male"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRII. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

BASE COUNT 167 a 204 c 63 g 127 t 9 others

ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 570;

Best Local Similarity 94.7%; Pred. No. 3.5e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TATTAGGCCACGCTCTC 20

Db 212 TATTAGGCCACGCTCTC 230

## RESULT 3

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

## FEATURES

## source

## Location/Qualifiers

## /organism="Trypanosoma brucei"

## /strain="TREU927/4 GUTat 10.1"

## /db\_xref="taxon:5691"

## /clone="Sheared DNA-57G10"

## /clone\_lib="Sheared DNA"

## /note="Vector: pUC18; Site\_1: SmaI; Constructed at the

## institute for Genomic Research (TIGR), Rockville, MD.

## Genomic DNA isolated from a cloned population of

## Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically

## sheared to give a tight size distribution (approx 2 kb).

## The v + i method used for the library construction is

## described in detail in Smith, H.O. and Venter, J.C.

## (Making small insert libraries for whole genome shotgun

## sequencing projects. In genome sequencing: A Practical

## Approach, eds. M. Vaubin and B. Barrell, Oxford University

## Press, 1999)."

## Seq primer: M13-Reverse

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

## Class: Shotgun.

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AATTACGCCACGCTCTC 20  
 |||||  
 Db 188 AATTACGCCACGCTCTC 169

## RESULT 4

B6429009/c

LOCUS 706 bp mRNA EST 14-MAR-2001  
 DEFINITION 602501922F1 NRH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4615444 5',  
 mRNA sequence.

ACCESSION

B6429009

VERSION

B6429009.1 GI:13335515

KEYWORDS

EST

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 706)

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

TITLE

NRH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov

Tissue

Tissue Procurement: CLONTECH Laboratories, Inc.

CDNA

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA

DNA Sequencing by: Incyte Genomics, Inc.

Clone

Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LNCM1368

row: 1 column: 05

High

quality sequence stop: 676.

Location/Qualifiers

1. 706

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4615444"

/lab\_host="NRH\_MGC\_75"

/lab\_host="PH10B (T1 phase-resistant)"

/note="Organ: Kidney; Vector: pNR-LIB (Clontech); Site: 1; SfiI (ggcgcctggcc); Site: 2; SfiI (ggcgcctggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCCATTATGCGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGCGCAGC-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NRH-MGC Library."

BASE COUNT

174 a

158 c

191 g

182 t

ORIGIN

1 others

Query Match

84.0%; Score 16.8; DB 12; Length 706;  
 Best Local Similarity 90.0%; Pred. No. 7.2e+02;

Matches

18; Conservative

0; Mismatches

2; Indels

0; Gaps

QY

1 AATTACGCCACGCTCTC 20  
 |||||

Db

41 AATTACGCCACGCTCTC 22  
 |||||

RESULT 5

AV755514/c

LOCUS 804 bp mRNA linear EST 19-OCT-2000  
 DEFINITION AV755514 BM Homo sapiens cDNA clone BMBD11 5', mRNA sequence.  
 ACCESSION AV755514  
 VERSION AV755514.1 GI:10913362  
 KEYWORDS EST  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 804)  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS

Gu, J., Zhao, M., Huang, Q., Xu, X., Li, Y., Peng, Y., Song, H., Xiao, H.,  
 Gu, Y., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Deng,  
 L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G.,  
 Yang, Y., Gao, G., Wang, Z., Zhang, Q., Chen, S., Han, Z., and Chen, Z.

TITLE

Homo sapiens cDNA BM clones

JOURNAL

Unpublished (2000)

COMMENT

Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919 (ex 45)  
 Fax: 86-21-50801922  
 Email: hanzg@chgc.sh.cn

This

clone is available at CHGC in Shanghai.

FEATURES

source

1. 804

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="BMBD11"

/lab\_host="BM"

/tissue\_type="Bone marrow"

/cell\_type="CD34+ Hematopoietic stem/progenitor cell"

/lab\_host="BM25.8"

/note="Vector: pTribex2; Site: 1: SfiI; Site: 2: SfiI"

BASE COUNT

177 a

189 c

221 g

212 t

ORIGIN

5 others

Query Match

84.0%; Score 16.8; DB 10; Length 804;  
 Best Local Similarity 90.0%; Pred. No. 7.5e+02;

Matches

18; Conservative

0; Mismatches

2; Indels

0; Gaps

QY

1 AATTACGCCACGCTCTC 20  
 |||||

Db

119 AATTACGCCACGCTCTC 100  
 |||||

RESULT 6

AA698952/c

LOCUS 219 bp mRNA linear EST 19-APR-2001  
 DEFINITION AA698952 5-prime HU Drosophila melanogaster head Bluescript  
 Drosophila melanogaster cDNA clone HU06045 5-prime, mRNA sequence.  
 ACCESSION AA698952  
 VERSION AA698952.1 GI:2701861  
 KEYWORDS EST  
 SOURCE fruit fly.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 219)  
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S., and Rubin, G.M.

TITLE

BDGP/HIMI Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: http://www.fruitfly.org/EST/estfruitfly.berkeley.edu  
 Plate: 60 row: D column: 9  
 High quality sequence stop: 157.  
 Location/Qualifiers  
 1. 219  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="HL06045"  
 /lab\_host="HU Drosophila melanogaster head Bluescript"  
 /sex="male and female"  
 /dev\_stage="adult"  
 /lab\_host="SOLR"

/note="Organ: head-brain & sensory organ; Vector: Bluescript SK; Site 1: EcoRI; Site 2: XhoI; Constructed using Stratagene ZAP-cDNA Synthesis kit. Oligo dt-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)"

## BASE COUNT

48 a 54 c 58 g 59 t

## ORIGIN

Query Match 82.0%; Score 16.4; DB 9; Length 219;  
Best Local Similarity 94.4%; Pred. No. 8.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTAACGCCACGCTCTC 20  
|||||

Db 96 ATTAACGCCACGCTCTC 79

## RESULT 7

LOCUS BH809948 294 bp DNA linear GSS 02-MAY-2002  
DEFINITION SALK\_036819 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK\_036819, DNA sequence.

ACCESSION BH809948

VERSION BH809948.1 GI:20387765

## KEYWORDS

## SOURCE

thale cress.

Arabidopsis thaliana

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Brassicales; Brassicaceae; Arabidopsis.

## REFERENCE

1 (bases 1 to 294)

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadlinab, C., Jeske, A., Karne, M., Kim, C.J., Parker, H., Predhals, L., Shinn, P., Zimmerman, J., and Ecker, J.R.

A sequence-indexed library of insertion mutations in the Arabidopsis Genome

unpublished (2001)

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA.

## FEATURES

Class: TDNA tagged.

Location/Qualifiers

1..294

/organism="Arabidopsis thaliana"

/strain="Columbia 0"

/db.xref="taxon:3702"

/clone="SALK\_036819"

/note="PCK was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

BASE COUNT

81 a 69 c 53 g 91 t

## ORIGIN

Query Match 82.0%; Score 16.4; DB 17; Length 294;

Best Local Similarity 94.4%; Pred. No. 9e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTAACGCCACGCTC 18  
|||||

Db 205 ATTAACGCCACGCTC 188

## RESULT 8

AA698954/c

LOCUS AA698954

DEFINITION HL06047 5prime HL Drosophila melanogaster head Bluescript

ACCESSION AA698954

VERSION AA698954.1 GI:2701883

KEYWORDS

EST.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Nephroletoidea; Drosophilidae; Drosophila.

1 (bases 1 to 307)

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S., and Rubin, G.M.

BDGP/HMI Drosophila EST Project

unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)

Plate: 60 row: D column: 11

High quality sequence stop: 219.

Location/Qualifiers

1..307

/organism="Drosophila melanogaster"

/db.xref="taxon:7227"

/clone="HL06047"

/clone.lib="HL Drosophila melanogaster head Bluescript"

/sex="male and female"

/dev.stage="adult"

/lab.host="SOLR"

/note="Organ: head-brain & sensory organ; Vector: Bluescript SK; Site 1: EcoRI; Site 2: XhoI. Constructed using Stratagene ZAP-cDNA Synthesis kit. Oligo dt-primed and directionally cloned at EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT

73 a 72 c 85 g 77 t

ORIGIN

Query Match 82.0%; Score 16.4; DB 9; Length 307;

Best Local Similarity 94.4%; Pred. No. 9.1e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTAACGCCACGCTCTC 20  
|||||

Db 96 ATTAACGCCACGCTCTC 79

## RESULT 9

LOCUS BI962036 573 bp mRNA linear EST 22-OCT-2001  
DEFINITION BI962036 1231 leafy spurge Lambda HybridAP 2.1 two-hybrid vector cDNA library Arabidopsis esula cDNA clone 16AR 5', mRNA sequence.

ACCESSION BI962036

VERSION BI962036.1 GI:16320239

KEYWORDS

EST.

leafy spurge.

Euphorbia esula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.

1 (bases 1 to 573)

Anderson, J.V., and Horvath, D.P.

Identification of RNAs expressed in underground adventitious buds of Euphorbia esula (leafy spurge)

unpublished (2000)

Contact: Anderson JV

Plants Science Research

USDA/ARS, Biosciences Research Lab

1605 Albrecht Blvd., PO Box 5674, Fargo, ND 58105, USA

Tel: 701 239 1263  
Fax: 701 239 1252  
Email: andersjvefargo.ars.usda.gov  
Seq primer: pbd5.

# FEATURES

## source

Location/Qualifiers

1..573  
/organism="Euphorbia esula"  
/db\_xref="taxon:3993"  
/clone="16AR"  
/clone.lib="leafy spurge lambda HybridAP 2.1 two-hybrid vector cDNA library"  
/tissue\_type="underground adventitious buds"  
/dev\_stage="3-day induced (decapitated)"

BASE COUNT 134 a 152 c 129 g 155 t 3 others

## ORIGIN

Query Match 82.0%; Score 16.4; DB 13; Length 573;  
Best Local Similarity 85.0%; Pred. No. 1.1e+03;  
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTACGCCACGCTCTC 20  
|||||  
Db 299 AGATTACGCCACGCTCTC 280

RESULT 10  
BML33610/c

LOCUS BML33610 679 bp mRNA linear EST 07-JAN-2002  
DEFINITION NXLV\_009\_F09\_F NXLV (Nsf Xylem Late wood Vertical) Pinus taeda cDNA  
clone NXLV\_009\_F09\_5', mRNA sequence.

ACCESSION BML33610  
VERSION BML33610.1 GI:17140689

KEYWORDS EST.  
SOURCE loblolly pine.

## ORGANISM

Pinus taeda  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

REFERENCE 1 (bases 1 to 679)  
AUTHORS Sederoff, R.  
TITLE Molecular Basis of Wood Formation in the Pine Megagenome  
JOURNAL Unpublished (2000)  
COMMENT Contact: Johnson, Arthur  
North Carolina State University  
Tel: 919 515 7800  
Fax: 919 515 7801  
Email: ajohnson@unity.ncsu.edu

FEATURES  
Seq primer: T3.  
Location/Qualifiers

1..679  
/organism="Pinus taeda"  
/strain="Coastal plain loblolly pine from North Carolina"  
/db\_xref="taxon:3352"  
/clone="NXLV\_009\_F09"  
/clone.lib="NXLV (Nsf Xylem Late wood Vertical)"  
/tissue\_type="primary xylem"  
/dev\_stage="late wood"  
/lab\_host="XLI-Blue"  
/note="Vector: pTRIPLEX; Site\_1: EcoRI; The library is from late (summer-August) wood, taken from below the crown of a 20 year old tree. The harvested xylem tissue was on the cusp between transitional and mature wood. NOTE: The sequences contain a 'cDNA adapter' between the EcoRI site and the start of the EST. The adapter sequence is AATTGGGCAATGAGCC."

BASE COUNT 177 a 101 c 174 g 210 t 17 others

## ORIGIN

Query Match 82.0%; Score 16.4; DB 13; Length 679;  
Best Local Similarity 89.5%; Pred. No. 1.1e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TATTACGCCACGCTCTC 20  
|||||

Db 558 TATTACGCCACGCTCTC 540

RESULT 11

LOCUS AA940956/c 700 bp mRNA linear EST 23-APR-2001  
DEFINITION LD23315.5prime LD Drosophila melanogaster embryo pot2 Drosophila melanogaster cDNA clone LD23315 5 similar to CG8240: FBan0008240  
/signal transduction' located on: X 16C1-16C1:: 04/10/2001, mRNA sequence.

ACCESSION AA940956  
VERSION AA940956.2 GI:13767616  
KEYWORDS EST.  
SOURCE fruit fly.

## ORGANISM

Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 700)  
AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.  
TITLE BSGP/HMT Drosophila EST Project  
JOURNAL Unpublished (2001)  
COMMENT On May 1, 1998 this sequence version replaced gi:3100869.  
Other ESTs: LD23315.3prime  
Contact: Stapleton, M.  
BDGP

Lawrence Berkeley National Lab  
One Cyclotron Rd  
Berkeley, CA 94720, USA  
Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
hit genomic AB003506: arm:x [11791364,17592357]  
estimated-cyto:16B4-16D6: 04/10/2001  
Plate: LD.233 row: B column: 3  
High quality sequence stop: 595  
POLYA-NO.

FEATURES  
Location/Qualifiers

1..700  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="LD23315"  
/clone.lib="LD Drosophila melanogaster embryo pot2"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="XLI Blue"  
/note="Organ: embryo; Vector: pOT2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2."

BASE COUNT 171 a 178 c 175 g 176 t

## ORIGIN

Query Match 82.0%; Score 16.4; DB 9; Length 700;  
Best Local Similarity 94.4%; Pred. No. 1.1e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTACGCCACGCTCTC 20  
|||||

Db 589 ATTACGCCACGCTCTC 572

RESULT 12  
LOCUS AA950399/c 872 bp mRNA linear EST 23-APR-2001  
DEFINITION LD28939.5prime LD Drosophila melanogaster embryo pot2 Drosophila melanogaster cDNA clone LD28939 5 similar to CG8240: FBan0008240  
/signal transduction' located on: X 16C1-16C1:: 04/10/2001, mRNA sequence.

ACCESSION AA950399  
VERSION AA950399.2 GI:13768163  
KEYWORDS EST.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 872)  
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G. M.  
 BDGP/HMTI Drosophila EST Project  
 Unpublished (2001)  
 On May 4, 1998 this sequence version replaced gi:3111771.  
 Other ESTs: LD28939.3prime  
 Contact: Stapleton, M.  
 BDGP

Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
 hit genomic AB003506: acm:x [11791364,11792357]  
 estimated-cyto:16B4-16D6: 04/10/2001  
 Plate: LD.289 row: D column: 3  
 High quality sequence stop: 739  
 POLYA-No.

## FEATURES

Location/Qualifiers  
 1..872

/organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"

/clone="LD28939"

/clone\_id="LD Drosophila melanogaster embryo pot2"

/sex="male and female"

/dev\_stage="0 to 24 hours mixed stage embryonic"

/lab\_host="XLI Blue"

/note="organ: embryo; Vector: pot2; Site:1; EcoRI: Site:2;  
 XhoI: Sized fractionated cDNAs were directly ligated into  
 pot2."

## BASE COUNT

220 a 224 c 224 g 204 t

## ORIGIN

Query Match

Best Local Similarity 94.0%; Score 16.4; DB 9; Length 872;  
 Pred. No. 1.2e+03;

Matches 17; Conservative

0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTAAGCCCGACGCTCTC 20

|||||

Db 592 ATTAAGCCCGACGCTCTC 575

## RESULT 13

AO752099/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1503 bp DNA linear GSS 19-JUN-1999  
 HS:5570..B2.G10.T7A.RPCT-11 Human Male BAC Library Homo sapiens  
 genomic clone Plate=1146 Col=20 Row=N, DNA sequence.  
 AO752099  
 AO752099.1 GI:5539257  
 GSS.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1503)  
 Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T.,  
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D. and  
 Hood, L.  
 Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 99380589  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCT-11. For BAC  
 library availability, please contact Pieter de Jong

(pieter@dejong.med.bufileo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering/bac.htm>)  
 or from Research Genetics (info@resgen.com). BAC end Web Server:  
<http://www.htsc.washington.edu>  
 Plate: 1146 row: N Column: 20  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence stop: 1503.

## FEATURES

Location/Qualifiers  
 1..1503

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"

/clone="Plate=1146 Col=20 Row=N"

/clone\_id="RPCT-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBAC3.6; Site:1: EcoRI; Site:2: EcoRI;  
 Male blood DNA was isolated from one randomly chosen donor  
 and partially digested with a combination of EcoRI and  
 EcoRI Methylase. Size selected DNA was cloned into the  
 pBAC3.6 vector at EcoRI sites"

## BASE COUNT

452 a 311 c 520 g 197 t 23 others

## ORIGIN

Query Match

Best Local Similarity 94.0%; Score 16.4; DB 17; Length 1503;  
 Pred. No. 1.4e+03;

Matches 17; Conservative

0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTAAGCCCGACGCTCTC 20

|||||

Db 206 ATTAAGCCCGACGCTCTC 189

## RESULT 14

CNS01MAP/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

940 bp DNA linear GSS 14-JUN-2001  
 Aopheles gambiae GSS T7 end of clone 21F09 of Notre-dame library  
 from strain Pest of Aopheles gambiae (African malaria mosquito),  
 genomic survey sequence.  
 A1150690  
 A1150690.1 GI:7011169  
 GSS.  
 African malaria mosquito.  
 Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 Anopheles.  
 1 (bases 1 to 940)  
 Genoscope.  
 Direct Submission  
 Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))  
 - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))  
 2 (bases 1 to 940)  
 Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissenbach, J.  
 Direct Submission  
 Submitted (16-FEB-2000) BMR, Institut Pasteur, 25, rue du Dr.  
 Roux, Paris 75015, France  
 This clone is from an A. gambiae BAC library provided by F.H.  
 Collins and sequenced by Genoscope in collaboration with the  
 Laboratory of Biochem. and Biol. Molec. of Insects, Institut  
 Pasteur.

## FEATURES

Location/Qualifiers  
 1..940

/organism="Anopheles gambiae"

/strain="Pest"

/db\_xref="taxon:7165"

/clone="21F09"

/clone\_id="Notredame1"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

/note="end : T7"

